In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institute shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

3/17/65
b

ADAPTIVE DECISION THRESHOLD RECEIVERS USING STOCHASTIC APPROXIMATION TECHNIQUES

A THESIS

Presented to

The Faculty of the Graduate Division

by

Maurice Joseph Bertrand Bouvier, Jr.

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

in the School of Electrical Engineering

Georgia Institute of Technology

January, 1968

ADAPTIVE DECISION THRESHOLD RECEIVERS USING

STOCHASTIC APPROXIMATION TECHNIQUES

Approved:

Chairman

Date approved by Chairman: 1/17/68

ACKNOWLEDGMENTS

It is a pleasure to express my sincere appreciation to Drs. R. H. Pettit, my thesis advisor, and A. M. Bush, reading committee member. This appreciation stems not only from their service as members of the thesis committee, but for many stimulating discussions, both prior to and during the research period. These informal discussions covered many aspects of statistical communication theory and made the course work and research considerably more meaningful and rewarding. I also wish to thank Dr. D. L. Finn for his service as a member of the reading committee.

Special thanks are due Dr. B. J. Dasher for his kind considerations, resulting in the financial support of a NDEA Fellowship.

I especially wish to acknowledge the personal assistance rendered by Russell B. Long, U. S. Senator from Louisiana, whose intervention in our behalf helped secure federal assistance following the flood of hurricane "BETSY" on September 10, 1965.

Finally, I wish to express my gratitude to my wife, Carolyn, without whose understanding and encouragement the successful completion of this program would not have been possible.

TABLE OF CONTENTS

Page

APPENDIX

LIST OF ILLUSTRATIONS

SUMMARY

Threshold type decision devices have an optimum setting which is defined by the performance criterion and the statistics of the input signals. Statistical decision theory provides the techniques for determining this optimum setting when complete statistical information is available. Often systems must be designed and built without the advantage of such complete knowledge or, as is more often the case, the statistics vary with time as a result of several uncontrollable factors. In such cases the receiver must be capable of learning, or adapting to, the optimum threshold setting while operating in the actual environment.

The mathematical theorems of stochastic approximation are used to derive receiver structures which exhibit such learning characteristics. Resulting receivers are trained using both the normal input, composed of a noise corrupted transmitted signal sequence, and the true signal sequence. With the threshold set at x_n, a decision is made regarding the signal transmitted during the nth signal interval, based on the input observed during that interval. A random variable, T(x), depending on the decision made, the true signal and the performance criterion, is generated. Using a decreasing sequence of positive numbers, $\{a_n\}$, as weighting factors, the threshold setting for the $n+1^{th}$ interval is defined by

$$x_{n+1} = x_n - a_n T(x) .$$

Under certain conditions, the iterative process converges to the optimum threshold, in some probabilistic sense, as the number of iterations becomes infinite. The recursive form of the receiver structures makes their synthesis and evaluation straightforward.

Receiver structures are obtained for eight, two signal, one threshold, performance criteria. Six of these criteria are such that the optimum threshold setting, x_o, is the solution of an equation of the form

$$N(x_o) = 0 .$$

For such criteria, the Robbins-Monro stochastic approximation method is directly applicable. Two criteria are such that the optimum threshold setting corresponds to the point at which a function achieves its minimum. For such criteria, the Robbins-Monro method is not so directly applicable and results in only an approximation to x_o. Receivers for these types of criteria can more readily be obtained using the Kiefer-Wolfowitz procedure. Although the Kiefer-Wolfowitz process converges to x_o exactly, its rate of convergence is less than that of the Robbins-Monro process.

General convergence properties are obtained for each receiver using the theorems of Sacks. These theorems establish the asymptotic distribution of $(x_n - x_o)$, that is, the probability distribution of the threshold error as the number of iterations becomes large. Under proper conditions, $(x_n - x_o)$ is shown to be asymptotically normally distributed with mean zero and a variance determined by the input statistics, performance criterion and number of iterations. The Robbins-Monro process is found to converge like $1/n$, whereas the Kiefer-Wolfowitz process converges like $1/n^{1/2}$. Convergence properties, determined from the asymptotic error variance, are investigated for uniform, gaussian and Ricean input signal statistics.

Receiver structures are obtained for two, three signal, two threshold, performance criteria. One criterion is such that the optimum threshold settings, denoted by the vector $\underline{x}_o$, are the solutions to the vector-valued equation

$$\underline{N}(\underline{x}_o) = \underline{0} .$$

The second is such that $\underline{x}_o$ corresponds to the vector for which the scalar-valued function, $R(\underline{x})$, attains its minimum. The first receiver structure is obtained using the multidimensional Robbins-Monro procedure; the second using the multidimensional Kiefer-Wolfowitz procedure. Sacks' multidimensional asymptotic distribution theorems are used to evaluate the convergence properties of these receivers.

Conditions under which the multidimensional theorems can be applied are more difficult to evaluate than are the one dimensional conditions. Resulting requirements on the input statistics are such that no general conclusions concerning the applicability of the theorems are obtained. For each criterion, one class of input statistics, for which the conditions are obviously satisfied, is determined.

CHAPTER I

INTRODUCTION

In binary communication systems the encoder selects one of two possible symbols, "0" or "1," for transmission during each signal interval. A rectangular pulse may be used as the transmitted waveform where the amplitude of the pulse takes on one of two possible levels depending on the symbol to be transmitted. The receiver makes an observation of the input waveform during each signal interval and decides which of the two symbols was selected. Similarly, in pulse radar or sonar systems, a rectangular pulse is transmitted and the receiver decides on the presence or absence of a reflected pulse. However, in all of these systems the decision making process is complicated by random noise which corrupts the transmitted waveforms. This noise will at times make one signal look like the other, so that an error is made in deciding which is present. The receiver input signal amplitude may then be described by an a posteriori probability density function, $p_o(V)$ or $p_1(V)$, depending on whether the "0" or "1" symbol, respectively, was selected. These density functions depend on both the amplitude of the received pulses and the statistical nature of the random noise.

A common method for deciding between two such received signals is to compare the received waveform with a decision threshold. Assuming the "1" symbol corresponds to the larger amplitude pulse, the decision rule is: if the observed value of the received waveform exceeds some value x, decide the "1" signal was transmitted during that interval; if less, decide the "0" signal was transmitted. Typical density functions and decision threshold are shown in Figure 1.

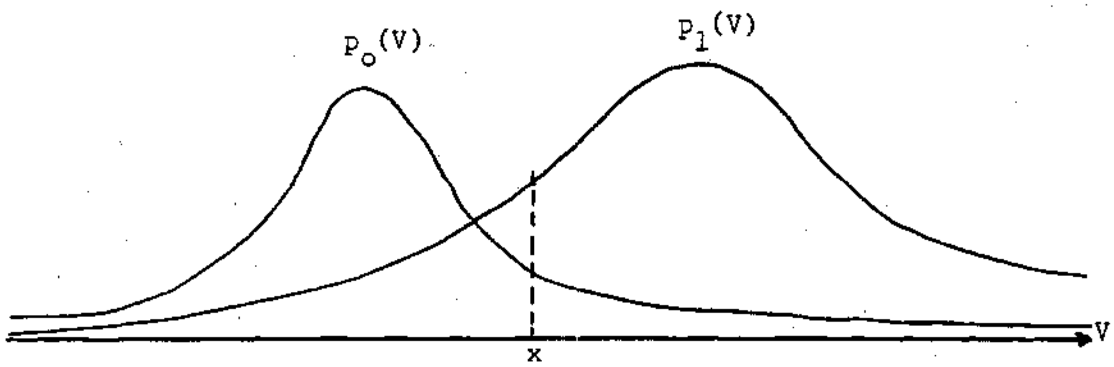


Figure 1. Typical A Posteriori Density Functions and Decision Threshold

There are four events which can occur during each decision interval, and each has a corresponding probability of occurrence given by

$$P[0|0] = P[\text{"0" decided given "0" transmitted}]$$

$$= \int_{-\infty}^{x} p_o(u)\, du$$

$$P[1|0] = P[\text{"1" decided given "0" transmitted}]$$

$$= \int_{x}^{\infty} p_o(u)\, du$$

$$P[1|1] = P[\text{"1" decided given "1" transmitted}]$$

$$= \int_{x}^{\infty} p_1(u)\, du$$

$$P[0|1] = P[\text{"0" decided given "1" transmitted}]$$

$$= \int_{-\infty}^{x} p_1(u)\, du$$

Two of these events represent decision errors which should be made as small as possible for best performance. However, if $p_o(V)$ and $p_1(V)$ overlap* changing the decision threshold to make the probability of one of the errors smaller in turn makes the probability of the other error larger. The relative importance of the four events will depend on the particular application. For example, in a radar system a high probability of falsely deciding a target is present may be acceptable, as long as the probability of missing a real target is low. However, if all defenses are wasted on false targets, recognizing a true target is then of only academic interest. Thus, for a particular application the relative importance of the four events is determined and used to define a criterion of "goodness of performance." One such criterion is that the average probability of error

*This is the only case of interest, since if $p_o(V)$ and $p_1(V)$ do not overlap, there exists a x such that a decision error is never made.

$$P(E) = P_o\, P[1|0] + P_1\, P[0|1] \qquad (1\text{-}1)$$

be minimum, where P_o and P_1 are the a priori probabilities of "0" and "1" symbols being transmitted, respectively.

Once a performance criterion has been selected, the *optimum threshold*, x_o, can be found in a straightforward manner if the signal levels, a priori probabilities and noise probability density functions are known. This problem is discussed in standard textbooks on decision or detection theory and the results are well known. For example, to minimize the average probability of error given by Equation (1-1), x_o is the solution to the equation

$$P_o\, p_o(x_o) = P_1\, p_1(x_o)\ . \qquad (1\text{-}2)$$

Given P_o, P_1, $p_o(V)$ and $p_1(V)$, x_o can be determined, and determined uniquely for most probability density functions of engineering importance.

However, in many practical applications one or more of the required quantities are not known exactly or may vary in some random manner with time. The signal levels may be unknown due to unknown RF signal attenuation between transmitter and receiver or may vary slowly due to changing weather conditions, equipment operating time or varying range. The a priori probabilities may not be known due to insufficient experimental data on the type of message transmitted and are also sure to change when the type of message is changed, say from speech to heartbeat. The noise probability density functions are never really known exactly and are certain to vary with time since they are functions of such external factors as interference from other transmitters and sky noise and such internal factors as receiver thermal noise and gain, all of which vary with time to some extent. Of the three, the noise density functions are the hardest to determine and often simply assumed. The assumption of gaussian noise is most often made because for many systems it is the only type which leads to mathematical solutions.

Under these conditions the optimum threshold cannot be determined directly. One approach may be to design the system to be optimum under the most probable or average conditions and accept the resulting inferior performance when the conditions are not "optimum." However, in some cases the degradation may be severe and in general this is not a very satisfying approach. An alternate approach is to design the system such that the unknown parameters can be learned. These learning or adaptive receivers must be capable of analyzing the input data and modifying their operating characteristics as the unknown parameters are learned. Learning may take place during a training period where the true signal sequence is known, supervised learning, or during a training period where the true signal sequence is unknown, nonsupervised learning. After sufficient training the receiver will be optimum or near optimum and is returned to normal operation. Obviously, system complexity and rate of learning or convergence to the optimum system are important considerations in the practical usefulness of such systems.

One class of adaptive receiver consists of the nonparametric detectors, so called because the functional form of the probability density functions need not be known. A survey of these techniques is given by Carlyle and Thomas (1) with more recent contributions widely available in the literature. These detectors usually do not attempt to learn the unknown parameters explicitly. Rather, during the supervised training period, they measure some dependent characteristic, such as the sample mean or the number of zero crossings, for both the signal conditions. During actual operation the same characteristic is measured in each decision interval, compared to the stored characteristic and a decision made. These techniques normally do not result in optimum receivers but perform near optimum for a large class of inputs. Their performance is usually superior to "optimum" systems operating in a "non-optimum" environment. However, these receivers are complex in that they require both calculation and storage capability and they provide little flexibility of performance criterion.

Abramson and Braverman (2) introduced both supervised and non-supervised learning techniques which have found wide application in pattern recognition systems. Although the nonsupervised systems have been more extensively studied, the approach is similar for supervised systems and the resulting complexity is of the same order of magnitude, as shown by Fralick (3). These methods usually assume that the forms of the input density functions are known, most often the gaussian density is used, with only a parameter such as the mean or variance being unknown. Some initial value for the unknown parameter is chosen and at each observation of the training sequence the density functions are modified. The amount of modification is determined by calculating a conditional density function, applying Bayes' Rule and integrating over one of the variables. After a sufficient training period an estimate of the unknown parameter is available but no knowledge of the a priori probabilities has been gained. If they are not known, some other method must be used to estimate them. Using these estimates of the a priori probabilities and unknown parameter, the optimum threshold is determined. Because only one parameter is estimated, convergence is relatively rapid. Hancock and Mix (4) applied these techniques to the decision threshold problem and include some convergence studies. However, the receivers are quite complex and appear to be unreasonable to implement for most input density functions.

Other nonsupervised adaptive techniques and applications are surveyed by Spragins (5). He points out that the receivers which converge to optimum receivers generally grow in size with the number of observations, n, as 2^n. If they are approximated by receivers with finite structures, they will not converge. These techniques generally result in very complex receiver operations also.

Stochastic approximation techniques provide an approach to adaptive threshold receivers which results in simple, realistic receiver structures. Resulting receivers do not attempt to first learn the unknown statistical parameters but learn the optimum threshold directly. This reduces the calculation and storage requirements considerably. The mathematical theorems were introduced by Robbins and Monro (6) in 1951 and have since been extensively studied by the mathematicians. Recently they have been applied to such diverse engineering problems as radar, matched filters, control systems and coding problems. Sakrison (7) gives a general discussion of the techniques and an excellent survey of possible applications. Kac (8) appears to have been the first to apply the general principles of stochastic approximation to adaptive threshold receivers. He suggested the receiver operations and proved convergence when the a priori probabilities are equal, and it is desired that the two probabilities of error be equal. Although his approach was slightly different from stochastic approximation and the receiver would not converge to the exact optimum, the principles were similar. Tong and Liu (9) determined the receiver operations when the receiver employs two thresholds, a null zone receiver, for two criteria of interest. Although they proved that the receivers converge as the training sequence becomes infinite, they, as Kac, made no theoretical investigation of the convergence rates and parameters which affect them. They did include some results of computer simulation studies. Cooper (10) presented the first detailed study of the convergence of adaptive threshold receivers developed with stochastic approximation methods. He considered a receiver to minimize the average probability of error when the a priori probabilities are equal. However, his work was primarily mathematical and provided little insight into the engineering aspects of the problem.

CHAPTER II

BASIC STOCHASTIC APPROXIMATION METHODS

Background

Consider first the problem of finding the point x_o at which a known function, $N(x)$, has a unique zero. If $N(x)$ is too complicated to find x_o by straightforward methods but is continuous in the neighborhood of x_o and has a positive derivative at x_o, a simple iterative process can be used to determine x_o. Some initial estimate of x_o, x_1, is chosen and $N(x_1)$ calculated. If $N(x_1)$ is positive, the next estimate, x_2, is selected such that x_2 is less than x_1; if $N(x_1)$ is negative, x_2 is selected greater than x_1. The rule for selecting x_{n+1} may be

$$x_{n+1} = x_n - a_n N(x_n) \qquad (2-1)$$

where a_n is a positive weighting factor. Note that the adjustment of x_n is always in the right direction, but may be either too large or too small. The weighting sequence may be decreasing so that the adjustment amount is damped out as n becomes large, but should form a divergent series in order to assure that any arbitrarily large initial error, $x_1 - x_o$, can be overcome. Using this process x_n approaches x_o as n becomes infinite.

This technique may be extended to the case where the function itself is not known exactly but a "noisy" estimate of $N(x)$, $T(x)$, can be observed for each x_n such that

$$E[T(x)] = N(x) . \tag{2-2}$$

Typical T(x) and N(x) are shown in Figure 2.

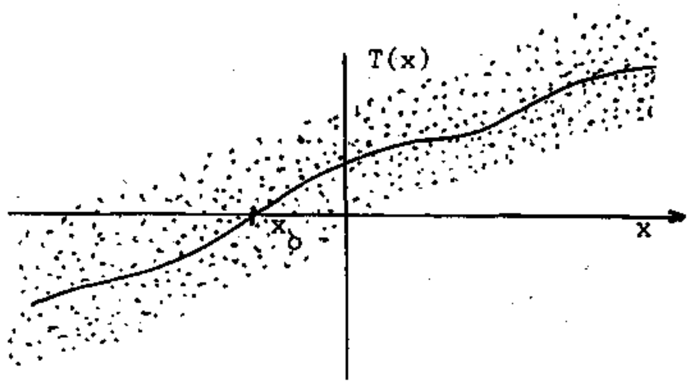


Figure 2. A "Noisy" Estimate of N(x)

One approach may then be to select x_1, make m independent observations of $T(x_1)$ as shown in Figure 3, estimate $N(x_1)$ by

$$\hat{N}(x_1) = \frac{1}{m} \sum_{i=1}^{m} T_i(x_1) \tag{2-3}$$

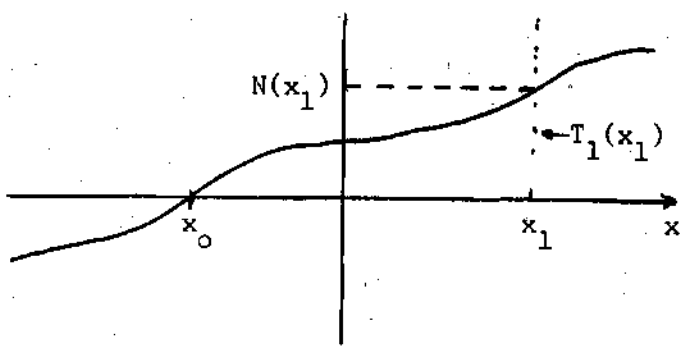


Figure 3. Obtaining an Estimate of $N(x_1)$

and select x_2 by

$$x_2 = x_1 - a_1 \hat{N}(x_1) .$$

The process is repeated using x_2 and m observations of $T(x_2)$ to estimate $N(x_2)$ and making a projection to x_3. In general adjustment is made according to the iterative rule

$$x_{n+1} = x_n - a_n \hat{N}(x_n) . \qquad (2\text{-}5)$$

Since only an estimate of $N(x_n)$ is used, the adjustment is sometimes in the wrong direction. But if m is large for each x_n, the average adjustment over a large number of iterations will be in the right direction and x_n would be expected to provide a good estimate of x_o as n becomes large.

The iterative methods of stochastic approximation employ this technique except that only one observation of $T(x_n)$ is used as an estimate of $N(x_n)$ so that

$$x_{n+1} = x_n - a_n T(x_n) . \qquad (2\text{-}6)$$

Since $T(x_n)$ is a crude estimate of $N(x_n)$ the adjustments are often in the wrong direction. However, the average or expected adjustment

$$E(x_{n+1} - x_n) = - a_n E[T(x_n)] \qquad (2\text{-}7)$$

$$= - a_n N(x_n)$$

is in the correct direction. If N(x) is an increasing function of x, as will often be the case in practical applications, the expected adjustment amount will be an increasing function of the magnitude of error. This operation is analogous to a servomechanism system where the larger the error, the stronger the "pull" to the correct setting, but movement in the wrong direction does occur occasionally. The mathematical theorems of stochastic approximation provide the conditions under which the process will converge to x_o, in some probabilistic sense, as n becomes infinite.

When locating the zero of a known function, more complicated methods, involving various order derivatives of $N(x_n)$, can be used to increase the rate of convergence to x_o. Likewise accelerated stochastic approximation methods have been described by Kesten (11) and Nikolic and Fu (12), among others, which help to overcome the major disadvantage of stochastic approximation, slow convergence. However, these methods also defeat the major advantage of stochastic approximation, simplicity, by requiring summation and information storage capabilities, and for this reason will not be considered.

<u>The Robbins and Monro Theorem</u>

Let N(x) be a fixed but unknown real-valued function of x and x_o be an unknown value of x such that

$$N'(x) \geq 0 \text{ for all } x \neq x_o , \tag{2-8}$$

$$N(x_o) = 0 \ , \tag{2-9}$$

and

$$N'(x_o) > 0 \ . \tag{2-10}$$

Define a random variable, $T(x)$, such that

$$E[T(x)] = N(x) \quad \text{for all } x \tag{2-11}$$

and for some finite, positive constant, C,

$$\Pr[|T(x)| \leq C] = 1 \quad \text{for all } x. \tag{2-12}$$

Define a sequence, $\{a_n\}$, such that

$$\sum_{n=1}^{\infty} a_n = \infty \tag{2-13a}$$

$$\sum_{n=1}^{\infty} a_n^{\,2} < \infty \ . \tag{2-13b}$$

Then the nonstationary Markov chain defined by the recursive equation

$$x_{n+1} = x_n - a_n T(x_n) \tag{2-14}$$

where x_1 is an arbitrary finite number, converges to x_o in mean square and with probability one as n approaches infinity.

This is the second theorem of Robbins and Monro's paper (6). It requires, by Equations (2-8), (2-9) and (2-10), that N(x) be nondecreasing, and that $N(x_o)$ and $N'(x_o)$ be defined. The first theorem of their paper, stated in Appendix I, relaxes these requirements by requiring only that N(x) be positive for finite values of x greater than x_o and negative for finite values of x less than x_o. The more restrictive theorem is sufficiently general for most applications and will be the one used where possible. However, for some criteria of interest, it will be more convenient to apply the less restrictive theorem. The conditions which affect the choice will be discussed when the situation arises.

Robbins and Monro proved mean square convergence of the process for both theorems in their paper. Blum (13) later proved convergence with probability one under equivalent conditions.

In applying the theorem to detection theory problems, the mathematical restrictions must be taken into account when formulating the criterion of performance. Then a means of properly generating T(x) must be found. The theorem places requirements on T(x) but does not suggest a method for defining it. Rather a candidate must be guessed and then tested to see if it satisfies the conditions. When the random variable is properly defined and all of the other conditions satisfied, the result is a recursive equation for successive threshold settings. The theorem assures, for any arbitrary initial threshold setting, that the equation converges to the optimum threshold setting as the training sequence becomes infinite. An adaptive threshold receiver based on Equation (2-14) is shown in Figure 4.

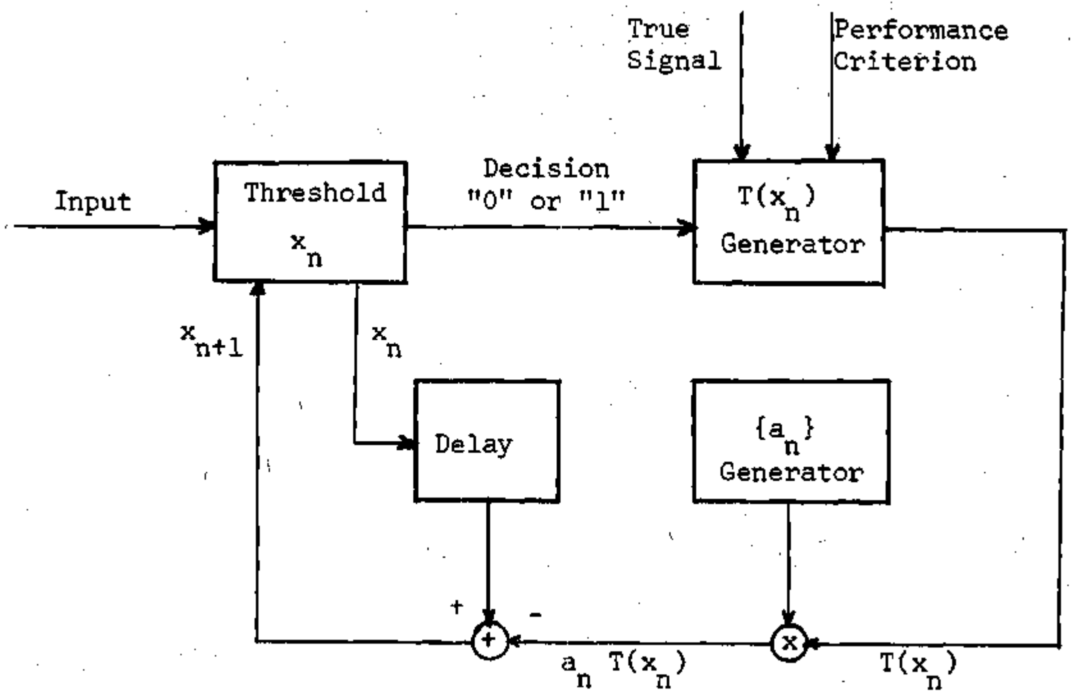


Figure 4. General Adaptive Threshold Receiver Block Diagram

During each signal interval the noise corrupted input signal is compared to the present threshold setting and a decision made regarding the symbol transmitted. This decision is compared to the true transmitted symbol and $T(x_n)$ generated according to a rule dependent on the performance criterion. $T(x_n)$ is then weighted by a_n to obtain the proper amount of threshold adjustment. This adjustment is subtracted from the present threshold setting, x_n, to provide the next threshold setting, x_{n+1}. No data storage or complex calculations are required; each input signal and threshold adjustment amount is used and discarded immediately. The receiver thus has the two desired properties: sure convergence and simplicity of implementation.

The learning is supervised in that the true signal sequence, as well as the noise corrupted signal sequence, must be known to the receiver during the training period. The receiver is also nonparametric, since the form of the probability density functions are not required, but is not the type detector normally considered in that class.

The random variable $T(x_n)$ must be generated such that its expectation determines a unique value of $N(x_n)$, independent of any particular sequence of $n - 1$ previously observed inputs. Since $T(x_n)$ is a function of only x_n and the observed input signal, this will be true if the input signals are statistically independent from interval to interval. In addition, $N(x_n)$ is independent of time so that the input statistics must be at least wide-sense stationary for the duration of the training period. It will be assumed throughout that the input signal satisfies these two conditions.

Returning to the statement of the theorem, the three conditions stated in Equations (2-8), (2-9) and (2-10) indicate the form in which the performance criterion must be expressed. Any function which has a unique zero and a defined derivative at the zero point can be expressed so that the conditions are satisfied. Equation (2-11) provides the basic requirement for defining $T(x)$. Equation (2-12) implicitly requires that the mean and variance of $T(x)$ be finite, since, with probability one, all of its possible values are equal to or less than some finite number. For all of the applications of interest, $T(x)$ will take on only finite values so that this condition will always be satisfied. Equation (2-13a) provides assurance that the process can overcome any magnitude of error introduced in the selection of the arbitrary starting point x_1. Equation (2-13b) implies the requirement that

$$\lim_{n \to \infty} a_n = 0 \tag{2-15}$$

so that as x_n converges to x_o, the amount that the threshold changes from its previous setting approaches zero. Thus this type of sequence allows any initial error to be overcome but also provides an increasingly more stable threshold setting as the setting approaches the optimum threshold.

One sequence which satisfies the conditions is

$$a_n = 1/n^p \tag{2-16}$$

where p is greater than one-half and less than or equal to one. Since one of the prime objectives is simplicity of implementation, the value of p will be taken as one unless other considerations dictate otherwise. This will require the generation of only integers to be used in the sequence $\{1/n\}$. Multiplying each term of a series by a positive, finite constant does not change the convergence properties of the series so that the sequence $\{A/n\}$ also satisfies the conditions. This may seem to be an added complication, but the parameter A is important in the studies of convergence rates to be considered later.

The Robbins-Monro theorem established the conditions under which stochastic approximation processes will converge as n becomes infinite. However, it provides no measure of how fast the processes will converge and in practice only a finite number of observations can be made. A measure of the convergence rate is needed in order that a stopping point, which will set the threshold within acceptable limits with high probability, can be selected. This measure is provided by the asymptotic distribution theorems of Sacks (14).

Sacks' First Theorem

Condition 1: $N(x)$ is a Borel-measurable function,

$$N(x_o) = 0 \tag{2-17}$$

and

$$(x - x_o)\, N(x) > 0 \tag{2-18}$$

for all finite values of x such that $x \neq x_o$.

Condition 2: For all x and some positive constant K,

$$|N(x)| \leq K|x - x_o| \, . \tag{2-19}$$

Condition 3: For all x

$$N(x) = \alpha_1 (x - x_o) + \delta(x,x_o) \tag{2-20}$$

where

$$\delta(x,x_o) = o(x - x_o) \tag{2-21}$$

as $|x - x_o| \to 0$ and where $\alpha_1 > 0$.

Condition 4: (a) $$\sup_{x} E\{[T(x) - N(x)]^2\} < \infty \tag{2-22}$$

(b) $$\lim_{x \to x_o} E\{[T(x) - N(x)]^2\} = \gamma^2 \tag{2-23}$$

Condition 5: For some $\varepsilon > 0$ and $v > 0$.

$$\sup_{|x - x_o| < \varepsilon} E\{|T(x) - N(x)|^{2+v}\} < \infty . \tag{2-24}$$

If the five conditions are satisfied and $a_n = A/n$ in Equation (2-14), where A is such that $A\,\alpha_1 > \frac{1}{2}$, $(x_n - x_o)$ is asymptotically normally distributed with mean zero and

$$\mathrm{Var}\,(x_n - x_o) = \frac{A^2\,\gamma^2}{n(2A\,\alpha_1 - 1)} . \tag{2-25}$$

Parzen (15) points out that one class of Borel-measurable functions consists of real functions which are continuous at all but possibly a finite number of points. For all problems of interest, N(x) is a member of this class. The other requirements of Condition 1 are satisfied by formulating the problem in accordance with either Equations (2-8), (2-9) and (2-10), or Equations (2-9), (A1-1) and (A1-2).

Condition 2 requires that N(x) remains bounded for values of x such that $|x - x_o|$ is finite but allows N(x) to become infinite as $|x - x_o|$ becomes infinite. For all realistic decision theory problems N(x) remains finite for all values of x, so that this condition is always satisfied.

Condition 3 requires that it be possible to consider N(x) as the sum of a straight line and a function, $\delta(x,x_o)$, which goes to zero faster than $|x - x_o|$ goes to zero. Then

$$\frac{d}{dx} \delta(x,x_o)\Big|_{x=x_o} = 0 \qquad (2\text{-}26)$$

and the condition is satisfied if N(x) is differentiable in a neighborhood of x_o and if N'(x) is continuous at x_o. Then

$$\alpha_1 = N'(x_o) \ , \qquad (2\text{-}27)$$

which is required to be positive by Equation (2-10).

Conditions 4(a) and 5 are always satisfied in decision theory problems because T(x) and its mean, N(x), will always be finite* so that all finite moments of T(x) will be finite. The limit in Condition 4(b), which exists in all the usual applications, defines γ, a parameter used in the asymptotic variance of the threshold error.

To summarize, Condition 1 can be satisfied if the basic process can be formulated; Conditions 2, 4(a) and 5 are always satisfied for practical adaptive threshold problems; Condition 3 is satisfied if the derivative of N(x) exists near x_o and is continuous at x_o and α_1 is defined as

$$\alpha_1 = N'(x_o) \ ;$$

and Condition 4(b) defines the parameter γ.

---

* Here, as elsewhere, we shall mean with a finite bound.

Using the definition of α_1 given by Equation (2-27), the error variance given by Equation (2-25) becomes

$$\text{Var } (x_n - x_o) = \frac{A^2 \gamma^2}{n[2AN'(x_o) - 1]} . \qquad (2\text{-}28)$$

Considering this expression a function of A, its minimum value occurs for

$$A = 1/N' (x_o) \qquad (2\text{-}29)$$

and is given by

$$\text{Min Var } (x_n - x_o) = \frac{\gamma^2}{n[N'(x_o)]^2} . \qquad (2\text{-}30)$$

If

$$A = K/N'(x_o) \qquad (2\text{-}31)$$

where K is greater than one-half, Equation (2-25) becomes

$$\text{Var } (x_n - x_o) = \frac{K^2}{2K - 1} [\text{Min Var } (x_n - x_o)] \qquad (2\text{-}32)$$

which is plotted in Figure 5. The advantage of a good choice of A is obvious. However, the value of A which minimizes the error variance depends on $N'(x_o)$ which in turn depends on the unknown probability

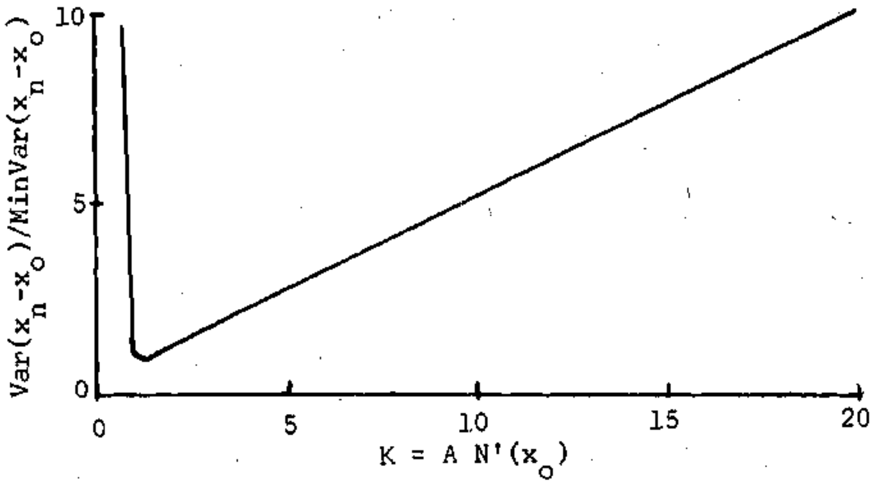


Figure 5. Influence of Parameter A on Convergence Rate

density functions. Cooper (16) suggests a method for estimating the optimum value of A, in effect adding a second adaptive section to the receiver, but it increases the complexity of the receiver considerably. It appears that good engineering judgment is the most reasonable solution.

Sacks' theorem provides the asymptotic distribution of $(x_n - x_o)$ only when

$$A\ N'(x_o) > \frac{1}{2}$$

and Equation (2-28) indicates that the error variance becomes infinite as

$$A\,N'(x_o) \rightarrow \frac{1}{2}\,.$$

However, the Robbins-Monro theorem assures convergence for any positive value of A. The convergence properties when

$$0 < A \leq 1/2N'(x_o)$$

are not known.

CHAPTER III

SINGLE THRESHOLD ADAPTIVE RECEIVERS

In this chapter stochastic approximation techniques are used to derive receiver structures for four criteria of engineering importance. The convergence rate is determined for each receiver using Sacks' theorems and evaluated for uniform, gaussian and Rayleigh input probability density functions.

Criterion: $P_o P[1|0] = \alpha$

In a particular application it may be necessary that the receiver operate such that the probability of transmitting the "0" symbol and deciding the "1" symbol is a predetermined value. The probability of this kind of error occurring, as a function of the threshold setting, x, is

$$\alpha(x) = P_o P[1|0] \tag{3-1}$$

$$= P_o \int_x^\infty p_o(u)\, du$$

The probability of the other kind of error, transmitting "1" and deciding "0," is given by

$$\beta(x) = P_1 \int_{-\infty}^{x} p_1(u)\, du \;. \tag{3-2}$$

Specifying a value of α determines the optimum threshold, x_o, which in turn fixes β. Hence, this criterion would be selected only when the error corresponding to α is very detrimental and any resulting value of β can be accepted.

Since $P[1|0]$ is equal to or less than one, α must be selected such that

$$P_o \geq \alpha \ ,$$

otherwise, no value of x will satisfy the criterion. In fact, selecting α equal to P_o would require that the system operate with

$$P[1|0] = 1 \ ,$$

which is not reasonable. When the a priori probabilities are not known, such a value of α may be inadvertently chosen. However, this is unlikely, since this criterion would normally be used only when α is required to be small.

Receiver Structure

After selecting an appropriate value of α, define

$$N(x) = \alpha - P_o \int_x^{\infty} p_o(u)\, du \ . \qquad (3\text{-}3)$$

Since the integral is a monotonically decreasing function of x, N(x) is a monotonically increasing function and the requirement of Equation

(2-8) is satisfied. Evaluating $N(x)$ at x_o gives

$$N(x_o) = \alpha - P_o \int_{x_o}^{\infty} p_o(u)\, du$$

$$= \alpha - \alpha = 0$$

so that Equation (2-9) is satisfied. Differentiating $N(x)$ with respect to x gives

$$\frac{dN(x)}{dx} = \frac{d\alpha}{dx} - P_o \frac{d}{dx} \int_{x}^{\infty} p_o(u)\, du$$

so that

$$N'(x) = P_o\, p_o(x) \tag{3-4}$$

and

$$N'(x_o) > 0$$

and Equation (2-10) is satisfied as long as $p_o(x_o)$ is not zero.

Define the discrete random variable

$$T(x) = \alpha - y_n \tag{3-5}$$

where

$$y_n = \begin{cases} 1 \text{ if "0" sent, "1" decided} \\ 0 \text{ if "0" sent, "0" decided} \\ 0 \text{ if "1" sent} \end{cases} . \tag{3-6}$$

Then

$$E[T(x)] = E(\alpha - y_n)$$

$$= \alpha - E[y_n]$$

$$= \alpha - P_o \int_x^\infty p_o(u)\, du$$

$$= N(x)$$

for all x and Equation (2-11) is satisfied. Since

$$0 \le |T(x)| \le 1$$

for all x, for any C equal to or greater than one,

$$\Pr\ [|T(x)| \le C] = 1$$

for all x and Equation (2-12) is satisfied. Let $\{a_n\}$ be any sequence which satisfies Equation (2-13). Then all of the conditions of the Robbins-Monro theorem are satisfied and the recursive process

$$x_{n+1} = x_n - a_n(\alpha - y_n) \tag{3-7}$$

converges to the optimum threshold, x_o, in mean square and with probability one as n becomes infinite. The adaptive receiver structure, based on Equation (3-7), is shown in Figure 6.

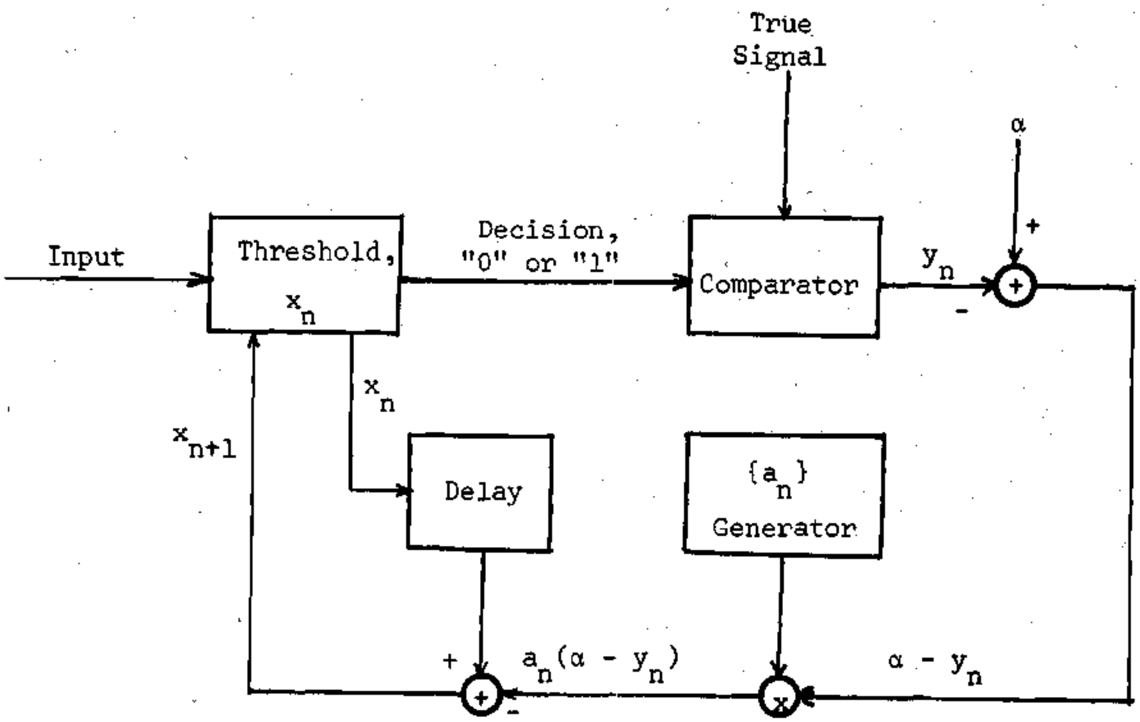


Figure 6. Adaptive Receiver for the Criterion: $P_o\ P[1|0] = \alpha$

Convergence

The Robbins-Monro theorem provides assurance that the adaptive threshold receiver of Figure 6 converges to the optimum threshold as the training sequence becomes infinite. Sacks' theorem discussed in Chapter II states that the asymptotic distribution, that is, the distribution for large n, of the threshold error is gaussian with mean zero and variance given by Equation (2-28). The number of iterations required to set the threshold within acceptable limits, with a given probability, can then be determined from this distribution. Therefore, the error variance can be considered a measure of the receiver's convergence rate. To apply the theorem, the engineering problem must be examined to determine whether or not the conditions are satisfied and, if they are, $N'(x_o)$ and γ^2 evaluated for use in Equation (2-28).

Condition 1 is satisfied by defining N(x) in accordance with Equation (3-3). Condition 2 is satisfied because

$$-1 \leq \alpha - P_o \leq N(x) \leq \alpha \leq 1$$

so that a K exists such that Equation (2-19) is satisfied. Using Equation (3-4) Condition 3 is satisfied with

$$\alpha_1 = P_o \, p_o(x_o) \; . \tag{3-8}$$

Since T(x) and N(x) are finite for all x, Conditions 4(a) and 5 are satisfied. Condition 4(b) is then used to determine γ^2 ,

$$\gamma^2 = \lim_{x \to x_o} E\{[T(x) - N(x)]^2\} \; .$$

With T(x) and N(x) continuous in the neighborhood of x_o and $N(x_o)$ zero, this becomes

$$\gamma^2 = \lim_{x \to x_o} E[T^2(x)] \ . \tag{3-9}$$

From the definition of T(x), Equation (3-5),

$$T^2(x) = \alpha^2 - 2\alpha y_n + y_n^{\ 2}$$

and using the definition of y_n, Equation (3-6),

$$y_n^{\ 2} = y_n$$

so that

$$T^2(x) = \alpha^2 - 2\alpha y_n + y_n \ .$$

Equation (3-9) then becomes

$$\gamma^2 = \lim_{x \to x_o} E[\alpha^2 - 2\alpha y_n + y_n]$$

$$= \lim_{x \to x_o} [\alpha^2 - 2\alpha P_o \int_x^\infty p_o(u)\, du + P_o \int_x^\infty p_o(u)\, du]$$

$$= \alpha^2 - 2\alpha P_o \int_{x_o}^\infty p_o(u)\, du + P_o \int_{x_o}^\infty p_o(u)\, du \ .$$

But

$$P_o \int_{x_o}^\infty p_o(u)\, du = \alpha$$

so that

$$\gamma^2 = \alpha^2 - 2\alpha^2 + \alpha$$

$$\gamma^2 = \alpha - \alpha^2$$

$$\gamma^2 = \alpha(1 - \alpha) \ . \qquad (3\text{-}10)$$

The theorem holds when a_n in Equation (3-7) and Figure 6 is A/n, where

$$A\ \alpha_1 > 1/2 \ .$$

Therefore, with

$$a_n = A/n$$

and

$$A > 1/2P_o\ p_o(x_o) \ ,$$

using Equations (3-8), (3-10) and (2-28), the error variance is given by

$$\mathrm{Var}\ (x_n - x_o) = \frac{1}{n}\ \frac{A^2\alpha(1 - \alpha)}{2AP_o p_o(x_o) - 1} \ . \qquad (3\text{-}11)$$

If A is selected to minimize this expression,

$$A = 1/N'(x_o) = 1/P_o\, p_o(x_o) \,,$$

and the minimum variance, Equation (2-30), becomes

$$\text{Min Var}\,(x_n - x_o) = \frac{1}{n}\;\frac{\alpha(1-\alpha)}{P_o^{\,2}\, p_o^{\,2}(x_o)} \,. \tag{3-12}$$

This expression indicates that the error variance tends to increase rapidly as $p_o(x_o)$ approaches zero. If α is such that x_o corresponds to a point on the probability density curve where the amplitude is small, such as the tail of a gaussian density, a long training time is required. This would be expected since in this region a normal threshold change will produce an almost unnoticeable change in the average number of decision errors. This effect will be quite apparent in the examples considered. The Min Var $(x_n - x_o)$ expression will be evaluated, rather than the general expression, in order to remove the dependence on A. The results should thus be considered lower bounds for convergence rates obtainable in practice.

If the input signal is uniformly distributed from -b to +b whenever the "0" signal is transmitted, the input is described by a uniform density function

$$p_o(V) = \begin{cases} \frac{1}{2b} \text{ for } |V| \le b \\ 0 \text{ for } |V| > b \end{cases} \tag{3-13}$$

$$m_V = E[V] = 0$$

$$\text{Var } V = E[(V - m_V)^2] = b^2/3 \ .$$

Then

$$\alpha = P_o \int_{x_o}^{\infty} (1/2b)\ du$$

$$\alpha = \frac{(b - x_o)P_o}{2b} \tag{3-14}$$

or solving for the optimum threshold in terms of α ,

$$x_o = b(1 - \frac{2\alpha}{P_o}) \tag{3-15}$$

and

$$p_o^{\ 2}(x_o) = \frac{1}{4b^2}$$

so that Equation (3-12) becomes

$$\text{Min Var } (x_n - x_o) = \frac{4b^2\alpha(1 - \alpha)}{n\ P_o^{\ 2}} \ . \tag{3-16}$$

The equation indicates that the error variance is directly proportional to the variance of the input signal. This is due to the effect discussed earlier; that is, as the variance increases, $p_o(x_o)$ decreases and the average number of decision errors becomes a less sensitive function of the threshold setting. The normalized optimum threshold, x_o/b, and normalized minimum error variance, Min Var $[(n/b^2)^{\frac{1}{2}}(x_n - x_o)]$, as functions of α are plotted in Figure 7 for various values of P_o.

Consider the performance when α is zero; the optimum threshold is

$$x_o \geq b \ .$$

Equation (3-16) indicates a zero training time but these convergence properties are for the asymptotic performance, that is, as the length of the training sequence becomes infinite. A zero error variance actually indicates a short training period. For α equal to zero, Equation (3-7) becomes

$$x_{n+1} = x_n + \frac{A}{n} y_n$$

where y_n is either zero or one. Thus, at each observation the threshold is either unchanged or moved up towards the region of optimum threshold, and for any initial setting the threshold is soon moved above b and the criterion satisfied. Similarly, when P_o and α are one, the optimum threshold is

$$x_o \leq -b$$

and Equation (3-7) becomes

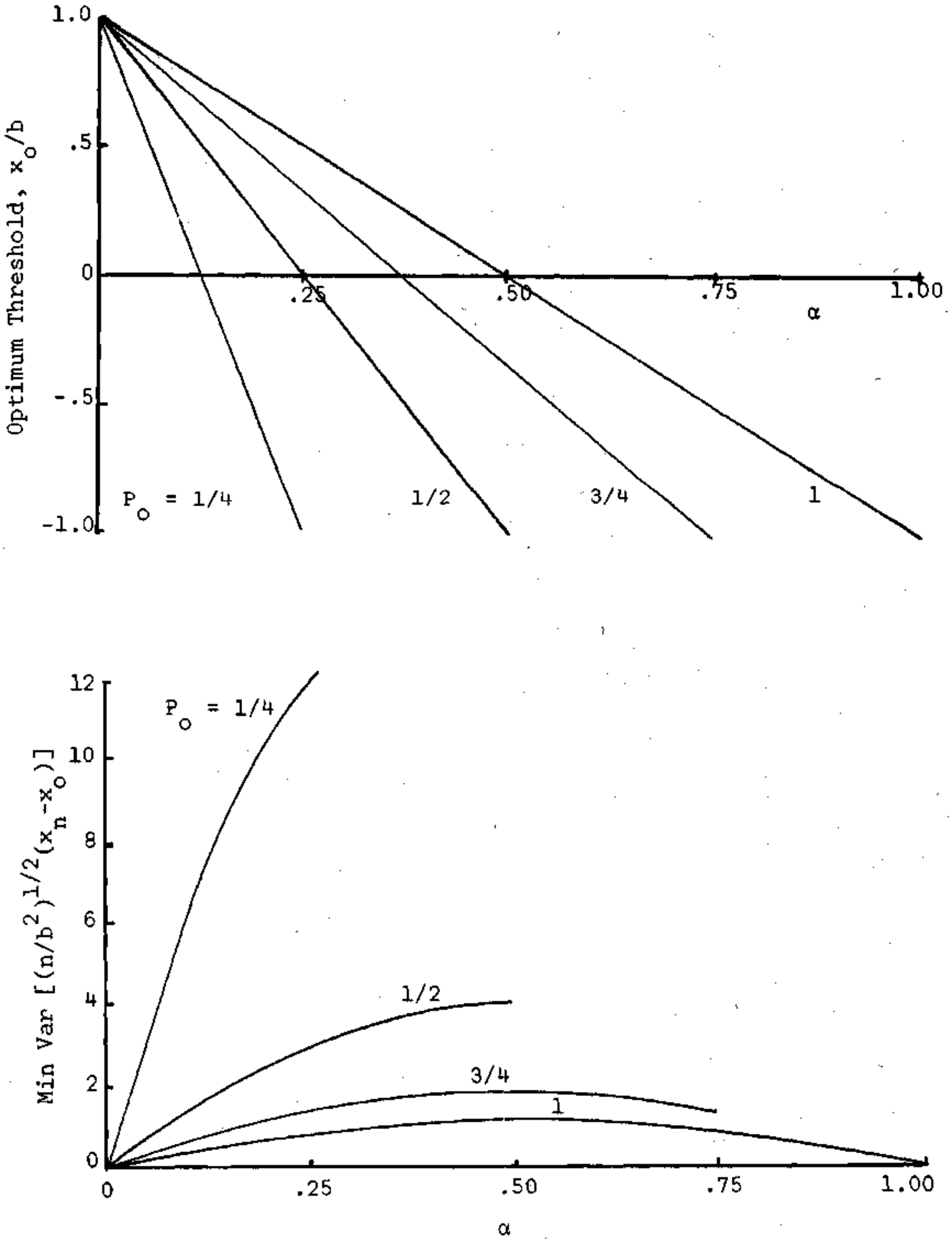


Figure 7. Optimum Threshold and Minimum Error Variance for Uniform Density and the Criterion: $P_o\ P[1|0] = \alpha$

$$x_{n+1} = x_n - \frac{A}{n}(1 - y_n) .$$

Adjustments are again always in the right direction, downward, and an error variance of zero is indicated. For all other values of α, an upward or downward adjustment is made after each observation. If at some point in the process the threshold is set at the optimum, it will be moved from optimum at the next iteration. Only the decreasing nature of a_n causes the system to become increasingly more stable and the error to approach zero as n becomes large. Thus, a non-zero error variance results for all reasonable values of α.

If the input signal has a zero mean, σ^2 variance gaussian distribution whenever the "0" symbol is transmitted, the input is statistically described by

$$p_o(V) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp \{- V^2/2\,\sigma^2\} \qquad (3\text{-}17)$$

$$m_V = E[V] = 0$$

$$\text{Var } V = \sigma^2 .$$

Then

$$\alpha = \frac{P_o}{\sqrt{2\pi}\,\sigma} \int_{x_o}^{\infty} \exp \{-V^2/2\sigma^2\}\, dV . \qquad (3\text{-}18)$$

Let

$$u = x/\sigma \ ,$$

and

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp \{-u^2/2\} \ ,$$

so that

$$\alpha = \frac{P_o}{\sqrt{2\pi}} \int_{u_o}^{\infty} \exp \{-V^2/2\} \, dV \tag{3-19}$$

and

$$p_o(x_o) = \frac{f(u_o)}{\sigma} \ . \tag{3-20}$$

Using standard tables of the gaussian distribution function, Equation (3-19) can be evaluated to plot the optimum threshold versus α, as shown in Figure 8.

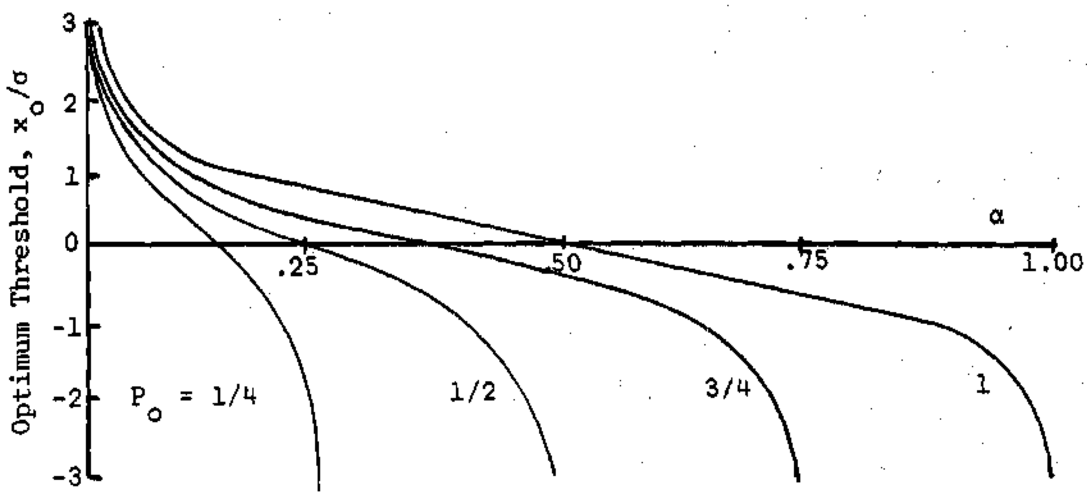


Figure 8. Optimum Threshold for Gaussian Density and the Criterion: $P_o P[1|0] = \alpha$

Using Equation (3-20), the minimum error variance becomes

$$\text{Min Var}\,(x_n - x_o) = \frac{\sigma^2}{n}\,\frac{\alpha(1-\alpha)}{P_o^{\;2}\, f^2(u_o)} \tag{3-21}$$

This expression is normalized by (σ^2/n), evaluated using standard tables and the result plotted in Figure 9 as a function of α.

The threshold error variance is again directly proportional to input signal variance. However, unlike the uniform density case, $f(u_o)$ is dependent on P_o so that the error variance is not exactly proportional to the inverse square of P_o, but is a more complicated function of P_o. For all values of P_o the error variance approaches infinity as α approaches P_o or zero. This would be expected since, as α approaches P_o or zero, the optimum threshold setting approaches minus or plus infinity, respectively. From any arbitrary initial setting, a large number of adjustments would normally be required to move the threshold to the neighborhood of the optimum value, even if every adjustment was in the correct direction. In addition, some adjustments are in the wrong direction, making convergence even slower.

As α approaches zero,

$$\alpha(1-\alpha) \cong \alpha$$

and Equation (3-21) becomes

$$\text{Min Var}\,(x_n - x_o) \cong \frac{\sigma^2}{n}\,\frac{\alpha}{P_o^{\;2}\, f^2(u_o)}\ . \tag{3-22}$$

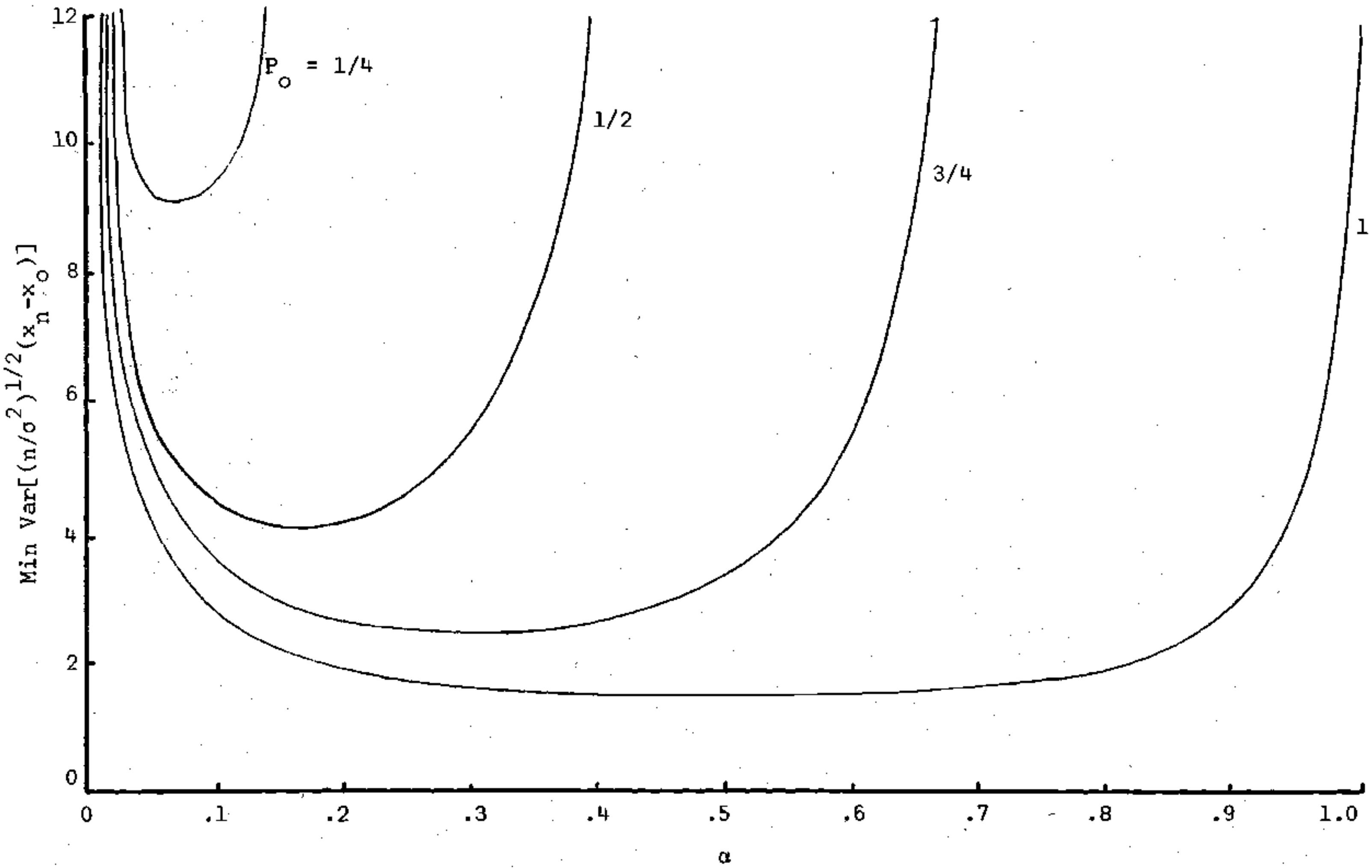


Figure 9. Minimum Error Variance for Gaussian Density and the Criterion: $P_o\ P[1|0] = \alpha$

When P_o is one and α approaches one,

$$\frac{\alpha(1 - \alpha)}{P_o^{\ 2}} \cong 1 - \alpha$$

and Equation (3-21) becomes

$$\text{Min Var } (x_n - x_o) \cong \frac{\sigma^2}{n} \frac{1 - \alpha}{f^2(u_o)} . \tag{3-23}$$

Because of the symmetry of the gaussian density function, the behavior of Equation (3-22) when P_o is one and α approaches zero is the same as the behavior of Equation (3-23) when α approaches one. Therefore, for P_o equal to one, the error variance curve is symmetrical about α equal to one-half. For all other values of P_o, as α approaches P_o,

$$\alpha(1 - \alpha) \cong C' > 0$$

and Equation (3-21) becomes

$$\text{Min Var } (x_n - x_o) \cong \frac{\sigma^2}{n} \frac{C'}{P_o^{\ 2}\ f^2(u_o)} . \tag{3-24}$$

The denominator goes to zero at the same rate as before, but the numerator is constant. Therefore for values of P_o other than one, the curves are not symmetrical about α equal to $P_o/2$, increasing faster as α approaches P_o than as α approaches zero.

Many receivers employ a narrow band filter followed by an envelope detector before the decision threshold. If the filter noise output is a sample function from a zero mean, σ^2 variance gaussian random process whenever the "0" symbol is transmitted, the threshold input is characterized by the Rayleigh probability density function,

$$p_o(V) = \begin{cases} \dfrac{V}{\sigma^2} \exp\{-V^2/2\sigma^2\} & \text{for } V \geq 0 \\ 0 & \text{for } V < 0 \end{cases} \tag{3-25}$$

$$m_V = \sigma\sqrt{\frac{\pi}{2}}$$

$$\mathrm{Var}\ V = (2 - \frac{\pi}{2})\sigma^2 \ .$$

Then

$$\alpha = P_o \int_{x_o}^{\infty} \frac{V}{\sigma^2} \exp\{-V^2/2\sigma^2\}\, dV$$

or after integrating,

$$\alpha = P_o \exp\{- x_o{}^2/2\sigma^2\} \ . \tag{3-26}$$

Letting

$$u_o = \frac{x_o}{\sqrt{2}\,\sigma}$$

$$\alpha = P_o \exp\{-u_o^2\} \tag{3-27}$$

or upon rearranging terms and taking the natural logarithm,

$$u_o^2 = \ln \frac{P_o}{\alpha}, \tag{3-28}$$

and the probability density function becomes

$$p_o(x_o) = \begin{cases} \dfrac{\sqrt{2}\, u_o}{\sigma} \exp\{-u_o^2\} & \text{for } u_o \geq 0 \\ 0 & \text{for } u_o < 0 \end{cases} \tag{3-29}$$

so that

$$p_o^2(x_o) = \begin{cases} \dfrac{2\, u_o^2}{\sigma^2} \exp\{-2u_o^2\} & \text{for } u_o \geq 0 \\ 0 & \text{for } u_o < 0 \end{cases}$$

or using Equation (3-28),

$$p_o^2(x_o) = \begin{cases} \dfrac{2 \ln (P_o/\alpha)}{\sigma^2\, P_o^2}\, \alpha^2 & \text{for } 0 \leq \alpha \leq P_o \\ 0 & \text{otherwise.} \end{cases} \tag{3-30}$$

Equation (3-28) is used to obtain the variation of the normalized optimum threshold as a function of α as shown in Figure 10.

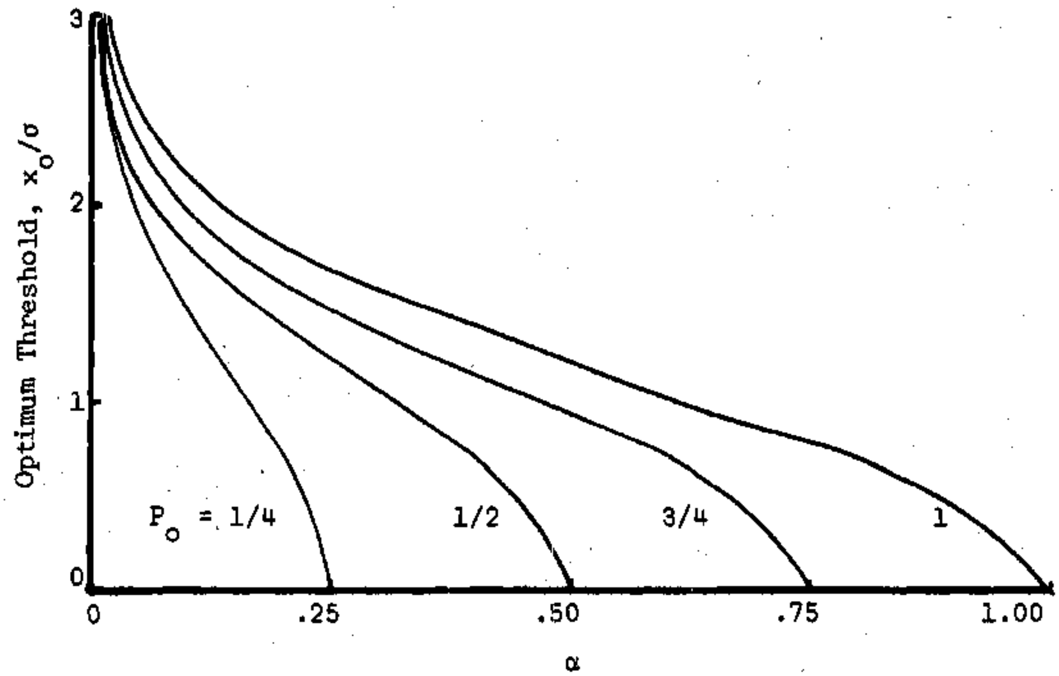


Figure 10. Optimum Threshold for Rayleigh Density and the Criterion: $P_o\ P[1|0] = \alpha$

Using Equation (3-30) in Equation (3-12), the minimum error variance becomes

$$\text{Min Var}\ (x_n - x_o) = \frac{\sigma^2}{n}\,\frac{1-\alpha}{2\alpha \ln(P_o/\alpha)}\ . \qquad (3\text{-}31)$$

This expression is normalized by (σ^2/n) and plotted as a function of α in Figure 11. The error variance has an unusual dependence on P_o; decreasing P_o still increases the variance but not at all like the inverse square relation for the uniform density case. However, the direct proportionality between error variance and input signal variance is still present.

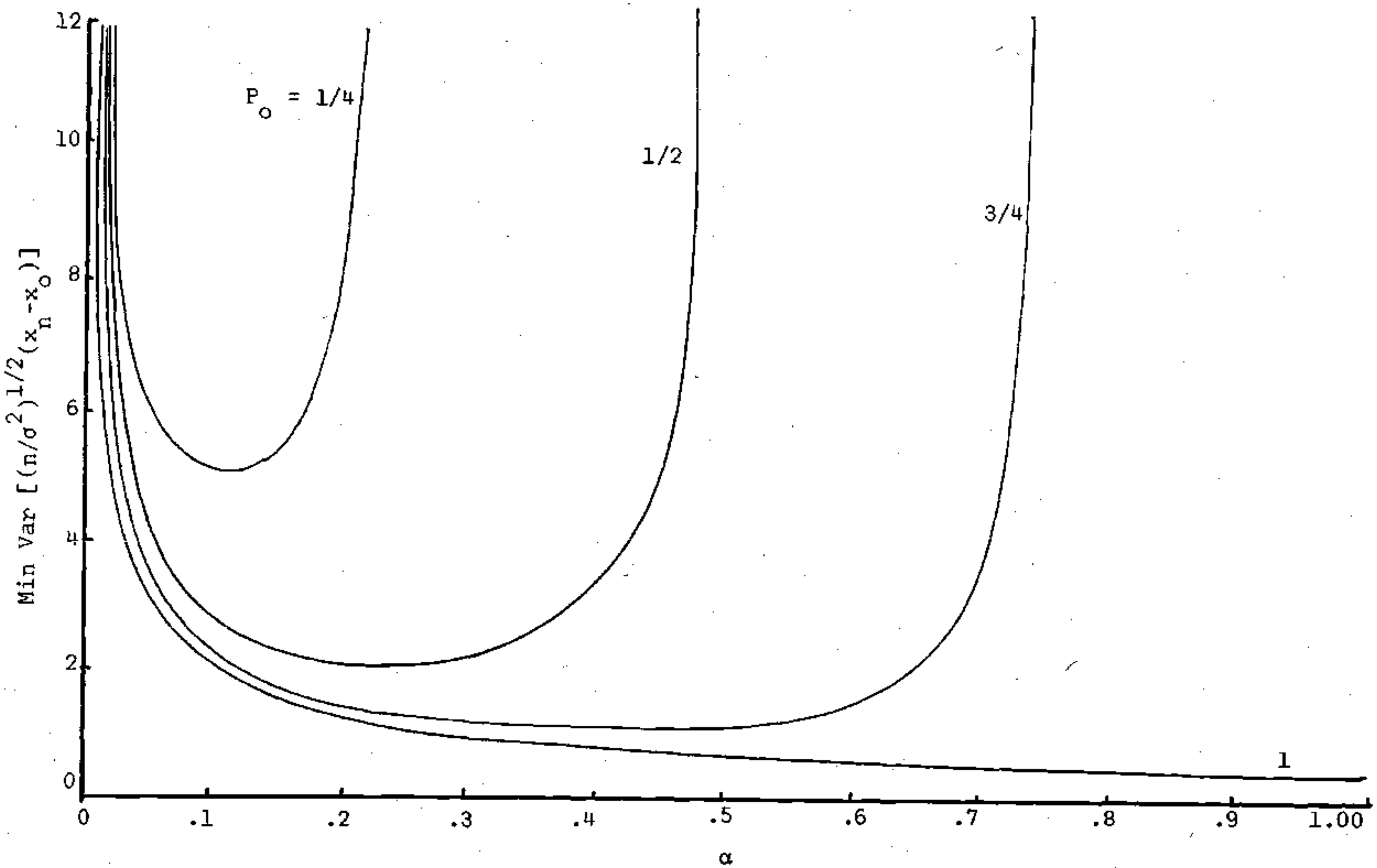


Figure 11. Minimum Error Variance for Rayleigh Density and the Criterion: $P_o \; P[1|0] = \alpha$

When α is very close to zero, the numerator is approximately one and

$$\alpha \ln P_o \cong 0$$

so that Equation (3-31) becomes

$$\text{Min Var } (x_n - x_o) \cong \frac{\sigma^2}{n} \frac{1}{2\alpha\ln(1/\alpha)}, \tag{3-32}$$

which goes to infinity as α approaches zero. Using Equations (3-27) and (3-28), Equation (3-31) can be rewritten as

$$\text{Min Var } (x_n - x_o) = \frac{\sigma^2}{n} \frac{1}{2\, u_o^2\, P_o} [\exp\, \{u_o^2\} - P_o] \,.$$

Using the series expansion

$$\exp\, \{y\} = 1 + y + \frac{y^2}{2!} + \frac{y^3}{3!} + \cdots$$

this becomes

$$\text{Min Var } (x_n - x_o) = \frac{\sigma^2}{n} \frac{1}{2u_o^2\, P_o} [1 + u_o^2 + \frac{u_o^4}{2!} + \frac{u_o^6}{3!} + \cdots - P_o]$$

$$= \frac{\sigma^2}{n} \frac{1}{2P_o} [\frac{1 - P_o}{u_o^2} + 1 + \frac{u_o^2}{2!} + \frac{u_o^4}{3!} + \cdots] \,.$$

As α approaches P_o, u_o approaches zero so that the higher order terms may be neglected and, using u_o^2 given by Equation (3-28),

$$\text{Min Var }(x_n - x_o) = \frac{\sigma^2}{n} \frac{1}{2P_o} \left[\frac{1 - P_o}{\ln(P_o/\alpha)} + 1\right] . \qquad (3\text{-}33)$$

For P_o other than one, as α approaches P_o the denominator approaches zero and the error variance goes to infinity. The rate of increase is less than that of Equation (3-32) as α approaches zero, so that the curves are not symmetrical. For P_o equal to one the first term in the brackets is zero and as α approaches P_o Equation (3-33) becomes

$$\text{Min Var }(x_n - x_o) = \frac{1}{2} \frac{\sigma^2}{n} . \qquad (3\text{-}34)$$

Operating with P_o equal to one implies a data sequence of only "0"s which would transfer no information and would not be of practical interest. Therefore, it should be assumed in practice that the error variance always becomes large as α approaches either P_o or zero.

The increasing nature of the error variance as α approaches zero arises for the same reason that the error variance for a gaussian input increases as α approaches P_o or zero; the optimum threshold setting becomes infinite. The performance as α approaches P_o is somewhat unexpected, since the optimum threshold is finite and performance would be expected to be similar to the uniform density case. The difference is that the uniform density has a step at -b whereas the Rayleigh density has a finite slope near zero. This causes the probabilities of upward and downward adjustments to change more slowly with x_n, resulting in a larger variation around the optimum threshold and the greatly increased minimum error variance.

<u>Criterion: $P[1|0] = \alpha'$</u>

In a particular application it may be necessary that the receiver operate such that the probability of deciding the "1" symbol was transmitted when in fact the "0" symbol was transmitted is a predetermined value. This probability, as a function of the threshold setting, x, is

$$\alpha'(x) = P[1|0]$$

$$\alpha'(x) = \int_{x}^{\infty} p_o(V)\, dV \ . \tag{3-35}$$

This criterion is important in systems, such as radar, where the a priori probabilities cannot be meaningfully defined so that criteria involving them are not practical. Rather, the probability of falsely detecting a target, or false alarm rate, is a useful criterion. Given the desired value of α', the optimum threshold can readily be determined from

$$\alpha' = \int_{x_o}^{\infty} p_o(V)\, dV \ . \tag{3-36}$$

There exists a solution to this equation for any α' in the range from zero to one.

<u>Receiver Structure</u>

One of two methods of generating the random variable T(x) for this criterion can be used, and the choice will depend on whether or not the a priori probabilities are known.

$\underline{P_o \text{ Known}}$. Define the discrete random variable

$$T(x) = \alpha' - \frac{1}{P_o} y_n \tag{3-37}$$

where

$$y_n = \begin{cases} 1 \text{ if "0" sent, "1" decided} \\ 0 \text{ if otherwise.} \end{cases} \tag{3-38}$$

Then

$$N(x) = E[T(x)]$$

$$N(x) = \alpha' - \frac{1}{P_o} [P_o \int_x^\infty p_o(V)\, dV]$$

$$N(x) = \alpha' - \int_x^\infty p_o(V)\, dV\ . \tag{3-39}$$

The integral is monotonically decreasing so that N(x) is monotonically increasing and Equation (2-8) is satisfied. Evaluating N(x) at x_o,

$$N(x_o) = \alpha' - \int_{x_o}^\infty p_o(V)\, dV$$

$$= \alpha' - \alpha' = 0$$

and Equation (2-9) is satisfied. Differentiating N(x) with respect to x gives

$$N'(x) = p_o(x) \tag{3-40}$$

and Equation (2-10) is satisfied as long as $p_o(x_o)$ is not zero. Since $T(x)$ takes on only finite values, there exists a C such that Equation (2-12) is satisfied. Let $\{a_n\}$ be any sequence which satisfies Equation (2-13); then all of the conditions of the Robbins-Monro theorem are satisfied and the recursive process

$$x_{n+1} = x_n - a_n(\alpha' - \frac{1}{P_o} y_n) \tag{3-41}$$

converges to the optimum threshold in mean square and with probability one as n becomes infinite. The adaptive receiver structure, based on Equation (3-41), is shown in Figure 12. Note that P_o must be known, as assumed, since it is one of the inputs during the training period.

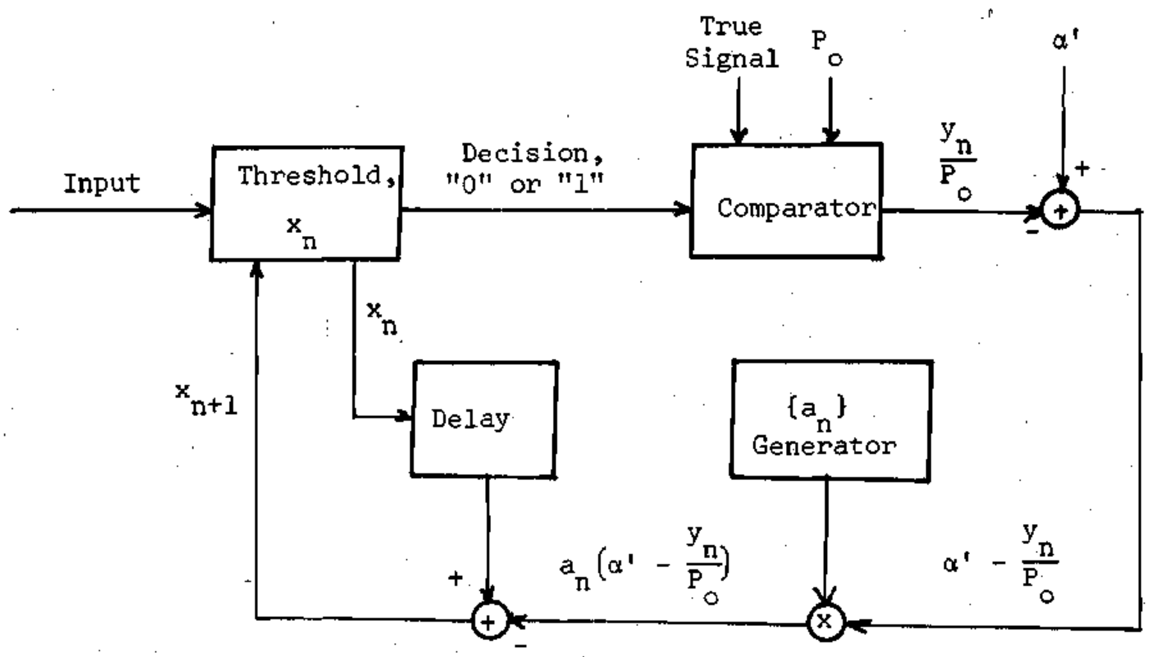


Figure 12. Adaptive Receiver for P_o Known and the Criterion: $P[1|0] = \alpha'$

<u>P_o Unknown</u>. Define the random variable

$$T(x) = z_n(\alpha' - y_n) \tag{3-42}$$

where

$$z_n = \begin{cases} 1 \text{ if "0" sent} \\ 0 \text{ if "1" sent} \end{cases} \tag{3-43}$$

and y_n is defined by Equation (3-38). Then

$$N(x) = E[T(x)]$$

$$N(x) = E[\alpha' z_n] - E[z_n y_n] \ .$$

But

$$z_n y_n = y_n$$

so that

$$N(x) = \alpha' P_o - P_o \int_x^\infty p_o(V)\, dV \tag{3-44}$$

and

$$N'(x) = P_o\, p_o(x) \ . \tag{3-45}$$

The requirements of Equations (2-8) and (2-9) are obviously satisfied, and Equation (2-10) is satisfied as long as $p_o(x_o)$ is not zero. Again T(x) takes on only finite values so that Equation (2-12) can be satisfied. Let $\{a_n\}$ be any sequence which satisfies Equation (2-13); then all of the conditions of the Robbins-Monro theorem are satisfied and the recursive process

$$x_{n+1} = x_n - a_n z_n(\alpha' - y_n) \tag{3-46}$$

converges to the optimum threshold as n becomes infinite. The receiver structure is shown in Figure 13.

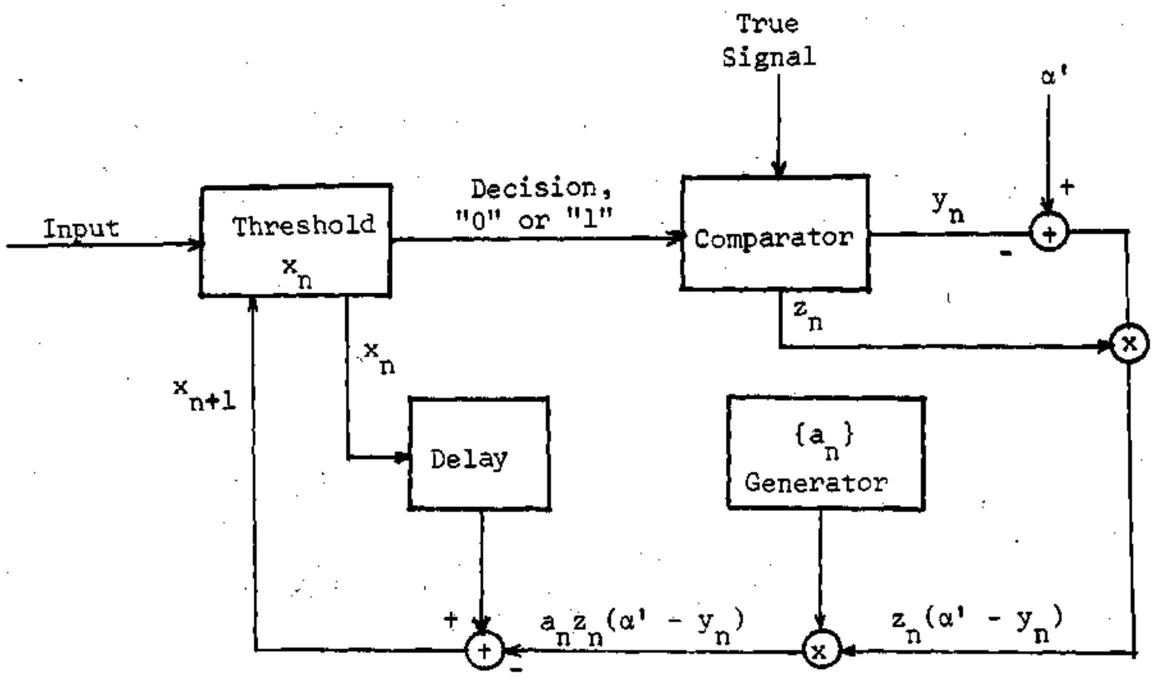


Figure 13. Adaptive Receiver for P_o Unknown and the Criterion: $P[1|0] = \alpha'$

Note that P_o need not be known since it does not appear as an input. The introduction of the random variable z_n has the effect of allowing the threshold to change only when the "0" symbol is transmitted. This removes the relative frequency of "0"s and "1"s from the adjustment part of the receiver.

For any criterion, the a priori probabilities appearing in T(x), N(x) and the error variance are the a priori probabilities of the symbols during the training sequence. The a priori probabilities appearing in the criterion itself are the a priori probabilities of the symbols during actual operation. To obtain a threshold setting during the training period that is optimum for operation, the two must obviously be the same. However, if the criterion of interest does not include P_o or P_1, any convenient values can be used during training. In fact, the criterion considered here does not depend on P_o, P_1 or $p_1(V)$. Therefore, a training sequence of all "0"s can be used. This is especially useful in a radar system where the amplitude of a target return, and consequently the shape of $p_1(V)$, is not a fixed value but depends on such factors as target range and size. The system can be trained on "noise only" and will still determine the optimum threshold for actual operation.

Convergence

Since T(x) and N'(x) depend on the method used to generate T(x), and these parameters enter into the convergence properties of a receiver, each receiver must be investigated separately.

P_o Known. Condition 1 of Sacks' theorem is satisfied by defining N(x) in accordance with Equation (3-39). Condition 2 is satisfied because $N(x)$ is finite for all x. Using Equation (3-40), Condition 3 is satisfied with

$$\alpha_1 = p_o(x_o) \ . \tag{3-47}$$

Since $T(x)$ and $N(x)$ are finite for all x, Conditions 4(a) and 5 are satisfied. To find γ^2,

$$T^2(x) = (\alpha' - \frac{1}{P_o} y_n)^2$$

$$T^2(x) = (\alpha')^2 - \frac{2\alpha'}{P_o} y_n + \frac{1}{P_o^{\,2}} y_n^{\,2}$$

so that, as in Equation (3-9),

$$\gamma^2 = \lim_{x \to x_o} E[T^2(x)]$$

$$\gamma^2 = \lim_{x \to x_o} E\{(\alpha')^2 - \frac{2\alpha'}{P_o} y_n + \frac{1}{P_o^{\,2}} y_n^{\,2}\}$$

$$\gamma^2 = \lim_{x \to x_o} \{(\alpha')^2 - \frac{2\alpha'}{P_o} [P_o \int_x^\infty p_o(V)\, dV] + \frac{1}{P_o^{\,2}} [P_o \int_x^\infty p_o(V)\, dV]\} \ .$$

But

$$\int_{x_o}^\infty p_o(V)\, dV = \alpha'$$

so that

$$\gamma^2 = (\alpha')^2 - 2\alpha'(\alpha') + \frac{1}{P_o}\alpha'$$

and after combining and rearranging terms,

$$\gamma^2 = \frac{\alpha'}{P_o}(1 - \alpha' P_o) \ . \tag{3-48}$$

The theorem holds when a_n in Equation (3-41) and Figure 12 is A/n, where, using Equation (3-47),

$$A > 1/2p_o(x_o) \ .$$

Then the error variance is

$$\mathrm{Var}\ (x_n - x_o) = \frac{1}{n}\ \frac{A^2\ \alpha'(1 - \alpha' P_o)}{P_o[2\ A\ p_o(x_o) - 1]} \ . \tag{3-49}$$

If A is selected to minimize this expression,

$$A = 1/p_o(x_o) \ ,$$

and the minimum variance is

$$\mathrm{Min\ Var}\ (x_n - x_o) = \frac{1}{n}\ \frac{\alpha'(1 - \alpha' P_o)}{P_o\ p_o{}^2(x_o)} \ . \tag{3-50}$$

Multiplying numerator and denominator by P_o and noting that

$$\alpha = \alpha' P_o$$

the expression becomes

$$\text{Min Var } (x_n - x_o) = \frac{1}{n} \frac{\alpha(1 - \alpha)}{P_o^{\,2}\, p_o^{\,2}(x_o)} \tag{3-51}$$

which is the same as the minimum error variance of the first criterion, given by Equation (3-12). To obtain minimum error variance versus α' curves based on Equation (3-50), it is only necessary to scale the α axis for the curves based on Equation (3-12) by $1/P_o$.

$\underline{P_o \text{ Unknown}}$. It is readily verified that Conditions 1, 2, 4(a) and 5 are satisfied by N(x) and T(x). Condition 3 is satisfied with

$$\alpha_1 = P_o\, p_o(x_o) \ , \tag{3-52}$$

as a result of Equation (3-45). From the definition of T(x), given by Equation (3-42),

$$T^2(x) = z_n^{\,2}[(\alpha')^2 - 2y_n\alpha' + y_n^{\,2}] \ .$$

But

$$z_n\, y_n = y_n \ ,$$

$$z_n^{\,2} = z_n \ ,$$

$$y_n{}^2 = y_n ,$$

so that

$$T^2(x) = z_n(\alpha')^2 - 2\, y_n\, \alpha' + y_n$$

and

$$E[T^2(x)] = P_o(\alpha')^2 - 2\alpha' P_o \int_x^\infty p_o(V)\, dV + P_o \int_x^\infty p_o(V)\, dV .$$

Proceeding as before,

$$\gamma^2 = P_o(\alpha')^2 - 2(\alpha')^2\, P_o + P_o \alpha'$$

and upon combining and rearranging terms,

$$\gamma^2 = P_o\, \alpha'(1 - \alpha') . \qquad (3\text{-}53)$$

Then when a_n in Equation (3-46) and Figure 13 is A/n, where

$$A > 1/2P_o\, p_o(x_o) ,$$

the error variance is

$$\mathrm{Var}\ (x_n - x_o) = \frac{1}{n}\ \frac{A^2\, P_o\, \alpha'(1 - \alpha')}{2\, A\, P_o\, p_o(x_o) - 1} . \qquad (3\text{-}54)$$

Selecting A to minimize this expression,

$$A = 1/P_o\, p_o(x_o),$$

results in

$$\text{Min Var } (x_n - x_o) = \frac{1}{n}\,\frac{\alpha'(1 - \alpha')}{P_o\, p_o^{\,2}(x_o)} \,. \tag{3-55}$$

Multiplying numerator and denominator by P_o gives

$$\text{Min Var } (x_n - x_o) = \frac{1}{n}\,\frac{\alpha(1 - \alpha')}{P_o^{\,2}\, p_o^{\,2}(x_o)} \,. \tag{3-56}$$

Since α is less than α',

$$1 - \alpha' < 1 - \alpha$$

so that, comparing Equations (3-51) and (3-56), convergence for the system with P_o known is slightly slower than convergence for the system with P_o unknown. Equation (3-46) indicates adjustment of the latter is made only when the "0" symbol is transmitted and the direction of adjustment depends on the receiver's decision. Equation (3-41) indicates adjustment of the former is also made when the "1" symbol is transmitted, but always in the negative direction. This extra movement, independent of the threshold setting and decision, increases the error variance.

If the input statistics are uniform, given by Equation (3-13), Equation (3-55) becomes

$$\text{Min Var } (x_n - x_o) = \frac{4b^2}{nP_o}\alpha'(1 - \alpha') . \tag{3-57}$$

The minimum error variance versus α' curve for this expression has the same shape as the P_o equal to one curve of Figure 7. Only the amplitude need be scaled by $1/P_o$ for the P_o of interest.

Equation (3-56) is

$$\frac{1 - \alpha'}{1 - \alpha} \tag{3-58}$$

times Equation (3-12), the minimum error variance for the criterion

$$P_o \, P[1|0] = \alpha .$$

Therefore, the error variance versus α' curves for the present criterion can readily be obtained from the earlier curves. The result when the input signal has a gaussian distribution is shown in Figure 14. The factor of Equation (3-58) approaches one as α' and α approach zero so that in the small α region, the curves are almost unchanged. As α' approaches one, α approaches P_o so that the factor of Equation (3-58) approaches zero. This behavior makes the nonsymmetrical curves of Figure 9, symmetrical for the present criterion, about α' equal to one-half.

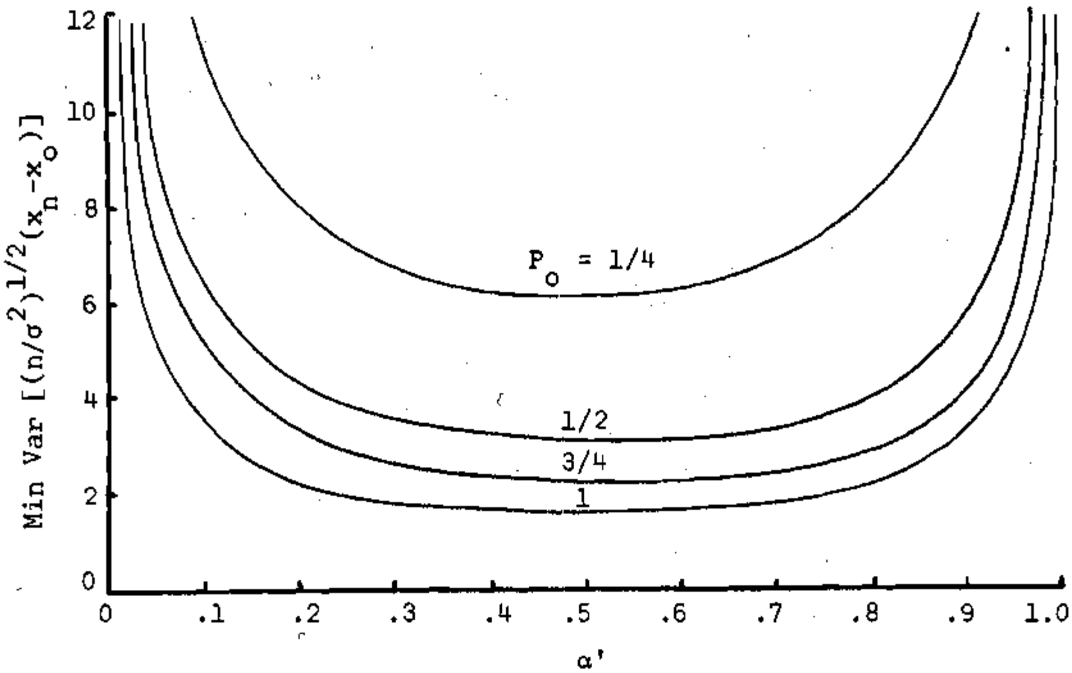


Figure 14. Minimum Error Variance for Gaussian Density, P_o Unknown and the Criterion: $P[1|0] = \alpha'$

The result when the input signal has a Rayleigh distribution is shown in Figure 15.

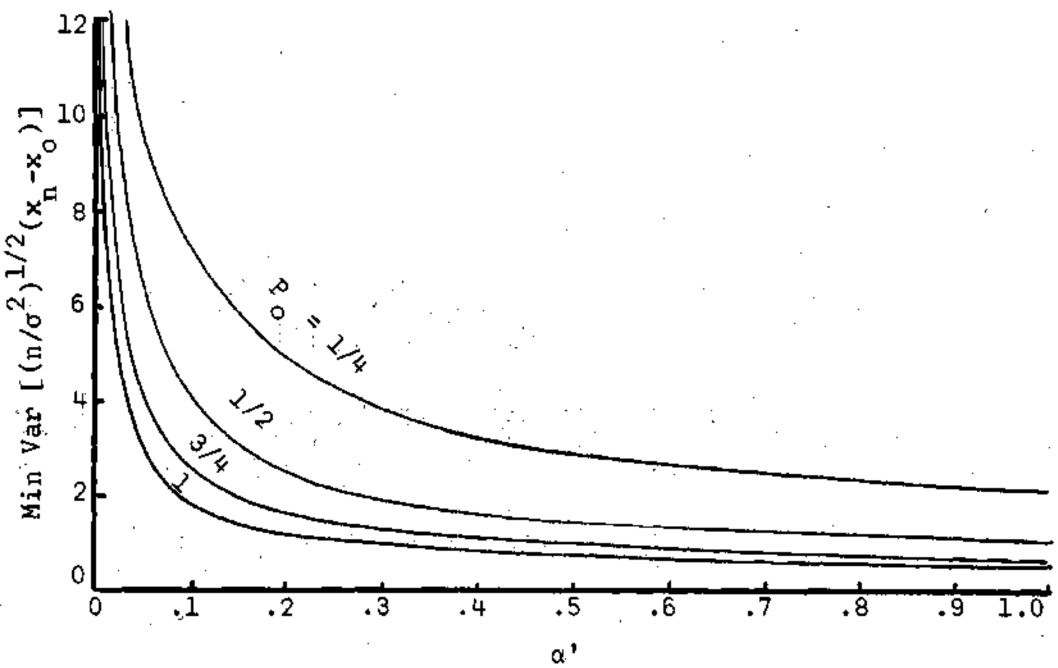


Figure 15. Minimum Error Variance for Rayleigh Density, P_o Unknown and the Criterion: $P[1|0] = \alpha'$

The curves, which increase without bound as α approaches P_o in Figure 11, approach a finite number for the present criterion. In terms of

$$u_o = \frac{x_o}{\sqrt{2}\,\sigma} ,$$

Equation (3-55) becomes for the Rayleigh density,

$$\text{Min Var } (x_n - x_o) = \frac{\sigma^2}{2nP_o} \frac{\exp\{u_o^2\} - 1}{u_o^2} . \tag{3-59}$$

As α' approaches one, u_o approaches zero so that

$$\lim_{\alpha' \to 1} \text{Min Var } (x_n - x_o) = \frac{\sigma^2}{2nP_o} . \tag{3-60}$$

Criterion: $P_1\, P[0|1] = \beta$

The first criterion considered was concerned with the probability of transmitting "0" and deciding "1." In some applications the probability of transmitting "1" and deciding "0" may be more important. This probability, as a function of the threshold setting, is

$$\beta(x) = P_1 \int_{-\infty}^{x} p_1(V)\, dV . \tag{3-61}$$

Specifying a value of β determines the optimum threshold, x_o, from the solution of

$$\beta = P_1 \int_{-\infty}^{x_o} p_1(V)\, dV , \tag{3-62}$$

as long as β is less than or equal to P_1.

Receiver Structure

Define the discrete random variable

$$T(x) = y_n - \beta \tag{3-63}$$

where

$$y_n = \begin{cases} 1 \text{ if "1" sent, "0" decided} \\ 0 \text{ if "1" sent, "1" decided} \\ 0 \text{ if "0" sent.} \end{cases} \tag{3-64}$$

Then

$$N(x) = E[T(x)] ,$$

$$N(x) = P_1 \int_{-\infty}^{x} p_1(V)\, dV - \beta . \tag{3-65}$$

The integral in this expression is monotonically increasing so that $N(x)$ is monotonically increasing. Evaluating $N(x)$ at x_o

$$N(x_o) = P_1 \int_{-\infty}^{x_o} p_1(V)\, dV - \beta$$

$$= \beta - \beta = 0 .$$

Differentiating N(x) gives

$$N'(x) = P_1\ p_1(x)\ . \qquad (3\text{-}66)$$

Therefore, Equations (2-8), (2-9), and (2-10) are satisfied as long as $p_1(x_o)$ is not zero. Equation (2-12) is satisfied because T(x) is finite for all x. Using an appropriate sequence, $\{a_n\}$, the process,

$$x_{n+1} = x_n - a_n\ (y_n - \beta)\ , \qquad (3\text{-}67)$$

converges to the optimum threshold, defined by Equation (3-62), as n approaches infinity. Comparing Equation (3-67) with Equation (3-7) indicates that the receiver for this criterion is essentially the same as the receiver of Figure 6.

Convergence

With

$$\alpha_1 = P_1\ p_1(x_o)\ , \qquad (3\text{-}68)$$

all of the conditions of Sacks' theorem are satisfied. Proceeding to find γ^2,

$$T^2(x) = (y_n - \beta)^2$$

$$T^2(x) = y_n^{\ 2} - 2\beta y_n + \beta^2\ ,$$

but

$$y_n^{\,2} = y_n$$

so that

$$\gamma^2 = \lim_{x \to x_\circ} E\{y_n - 2\beta y_n + \beta^2\}$$

$$\gamma^2 = \beta - 2\beta(\beta) + \beta^2$$

$$\gamma^2 = \beta(1 - \beta) \ . \tag{3-69}$$

Therefore with

$$a_n = A/n$$

and

$$A > 1/2P_1 p_1(x_\circ) \ ,$$

the error variance is

$$\mathrm{Var}\ (x_n - x_\circ) = \frac{1}{n} \ \frac{A^2\beta(1 - \beta)}{2\ A\ P_1\ p_1(x_o) - 1} \ . \tag{3-70}$$

For

$$A = 1/P_1 p_1(x_o) ,$$

this becomes

$$\text{Min Var } (x_n - x_o) = \frac{1}{n} \frac{\beta(1 - \beta)}{P_1^{\ 2} \ p_1^{\ 2}(x_o)} . \qquad (3\text{-}71)$$

These are of the same form as the corresponding expressions, Equations (3-11) and (3-12), for the criterion

$$P_o \ P[1|0] = \alpha .$$

Thus for symmetrical probability density functions, such as the uniform or gaussian densities, convergence properties are the same for these two criteria. However, the filter-envelope detector case is of interest and must be studied for each criterion separately.

If the received signal consists of a transmitted sine wave plus gaussian noise, the output of a narrowband filter will be

$$A_1 \cos\omega t + n(t) \qquad (3\text{-}72)$$

where n(t) is a sample function from a narrowband gaussian process, $N(0,\sigma^2)$. The envelope detector output, and decision threshold input, will have a Ricean probability density function given by

$$p_1(V) = \begin{cases} \dfrac{V}{\sigma^2} \exp\left[-\dfrac{V^2 + A_1^{\,2}}{2\sigma^2}\right] I_o\left(\dfrac{A_1 V}{\sigma^2}\right) & \text{for } V \geq 0 \\ 0 & \text{for } V < 0 \end{cases} \qquad (3\text{-}73)$$

where $I_o(x)$ is the modified Bessel function of the first kind of order zero. For this density function,

$$\beta = P_1 \int_0^{x_o} \frac{V}{\sigma^2} \exp\left[-\frac{V^2 + A_1^{\,2}}{2\sigma^2}\right] I_o\left(\frac{A_1 V}{\sigma^2}\right) dV . \qquad (3\text{-}74)$$

Let

$$v = V/\sigma ,$$

$$B = A_1/\sigma ,$$

$$u_o = x_o/\sigma ,$$

and

$$f_1(u_o) = u_o \exp\left[-\frac{u_o^{\,2} + B^2}{2}\right] I_o(Bu_o) , \qquad (3\text{-}75)$$

then

$$\beta = P_1 \int_0^{u_o} v \exp\left[-\frac{v^2 + B^2}{2}\right] I_o(Bv)\, dv \qquad (3\text{-}76)$$

and

$$\text{Min Var } (x_n - x_o) = \frac{\sigma^2}{n} \frac{\beta(1 - \beta)}{P_1^{\,2}\, f_1^{\,2}(u_o)} . \qquad (3\text{-}77)$$

β may be rewritten in terms of Marcum's Q Function (17)

$$Q(x,y) = \int_y^{\infty} t \exp\left[-\frac{t^2 + x^2}{2}\right] I_o(xt)\, dt , \qquad (3\text{-}78)$$

which also has been tabulated by Marcum (18), as

$$\beta = P_1 - P_1\, Q(B,u_o) . \qquad (3\text{-}79)$$

Using Equations (3-75) and (3-79), the minimum error variance is evaluated for two values of A_1/σ, and shown in Figure 16. The convergence time increases with A_1/σ due to the corresponding decrease in $p_1(x_o)$.

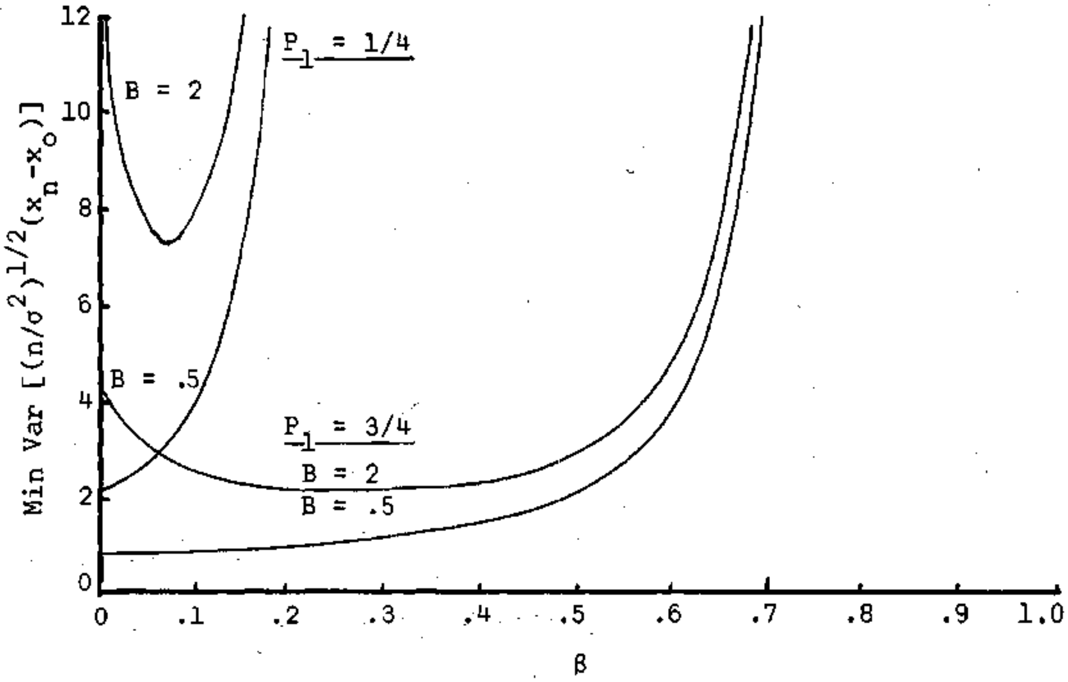


Figure 16. Minimum Error Variance for Ricean Density and the Criterion: $P_1 \; P[0|1] = \beta$

Criterion: $P[0|1] = \beta'$

The second criterion considered was concerned with the probability of deciding "1" given that a "0" was transmitted, and the importance of this criterion for radar-type systems was discussed. The analogous criterion, the probability of deciding "0" given that a "1" was transmitted, does not appear to have such wide application but may be of interest in some systems. This probability is

$$\beta'(x) = \int_{-\infty}^{x} p_1(V) \, dV \qquad (3\text{-}80)$$

so that by specifying a desired value of β', the optimum threshold can be determined from

$$\beta' = \int_{-\infty}^{x_0} p_1(V)\, dV \;. \tag{3-81}$$

Receiver Structure

Again there are two methods available for generating T(x) with the choice depending on whether or not the a priori probabilities are known. In fact, the similarities noted between the first and third criteria carry over to the second and fourth criteria. Therefore, only the pertinent definitions and results are included for this criterion.

$\underline{P_1 \text{ Known}}$. With

$$y_n = \begin{cases} 1 \text{ if "1" sent, "0" decided} \\ 0 \text{ if otherwise ,} \end{cases} \tag{3-82}$$

the iterative equation defining the threshold setting is

$$x_{n+1} = x_n - a_n\left(\frac{y_n}{P_1} - \beta'\right) , \tag{3-83}$$

which is similar to Equation (3-41).

$\underline{P_1 \text{ Unknown}}$. With y_n defined by Equation (3-82) and

$$z_n = \begin{cases} 1 \text{ if "1" sent} \\ 0 \text{ if "0" sent} \end{cases} \tag{3-84}$$

the iterative equation defining the threshold setting is

$$x_{n+1} = x_n - a_n\, z_n\, (y_n - \beta') , \tag{3-85}$$

which is analogous to Equation (3-46).

Convergence

$\underline{P_1 \text{ Known}}$. For

$$a_n = A/n ,$$

where

$$A > 1/2p_1(x_o) ,$$

$$\mathrm{Var}\ (x_n - x_o) = \frac{1}{n}\ \frac{A^2\ \beta'(1 - \beta' P_1)}{P_1[2\ A\ p_1(x_o) - 1]} \tag{3-86}$$

and if

$$A = 1/p_1(x_o) ,$$

$$\mathrm{Min\ Var}\ (x_n - x_o) = \frac{1}{n}\ \frac{\beta'(1 - \beta' P_1)}{P_1\ p_1^{\,2}(x_o)} . \tag{3-87}$$

Noting that

$$\beta = \beta' P_1 ,$$

this can be rewritten as the minimum error variance, Equation (3-71), for the criterion

$$P_1 \; P[0|1] = \beta \; .$$

Hence, the curves for the present criterion can be obtained from the earlier criterion by scaling the β axis by $1/P_1$.

<u>P_1 Unknown</u>. For

$$a_n = A/n \; ,$$

where

$$A > 1/2P_1 \; p_1(x_o) \; ,$$

$$\mathrm{Var}\;(x_n - x_o) = \frac{1}{n} \; \frac{A^2 \; P_1 \beta'(1 - \beta')}{2 \; A \; P_1 \; p_1(x_o) - 1} \tag{3-88}$$

and if

$$A = 1/P_1 p_1(x_o) \; ,$$

$$\mathrm{Min\;Var}\;(x_n - x_o) = \frac{1}{n} \; \frac{\beta'(1 - \beta')}{P_1 \; p_1{}^2(x_o)} \; . \tag{3-89}$$

When this expression is evaluated for the Ricean density, given by Equation (3-73), the result is as shown in Figure 17.

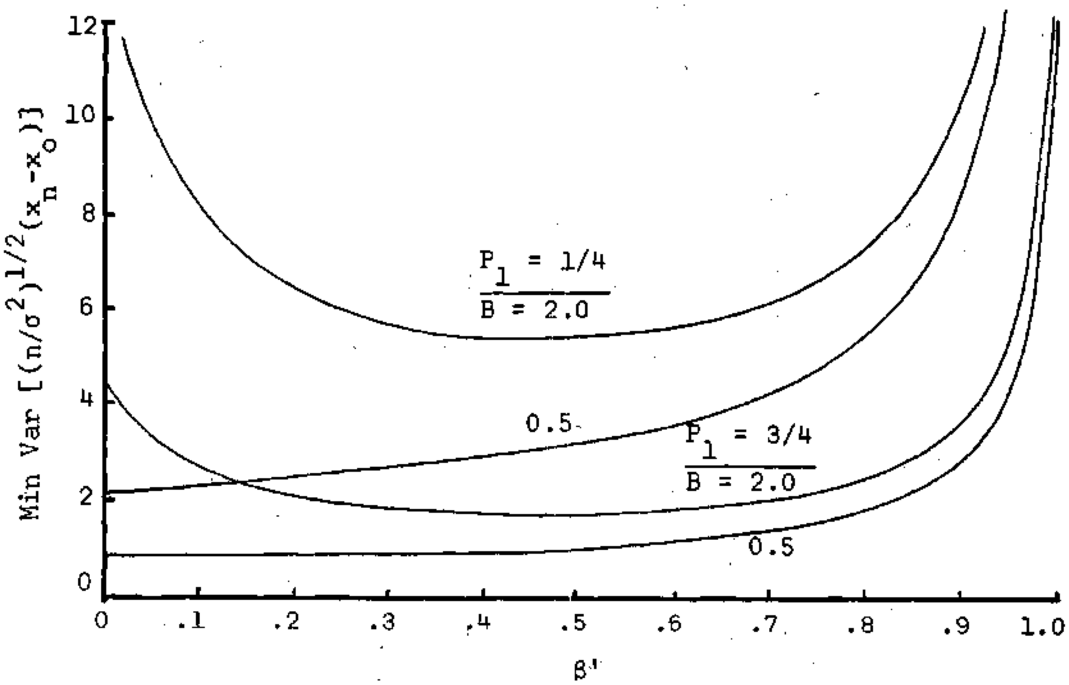


Figure 17. Minimum Error Variance for Ricean Density, P_1 Unknown and the Criterion: $P[0|1] = \beta'$

<u>Criterion: $P_o \, P[1|0] = P_1 \, P[0|1]$</u>

In many communication applications, performance is considered optimum when the two kinds of error have equal probability of occurrence. The optimum threshold for this criterion is defined by

$$P_o \int_{x_o}^{\infty} p_o(V)\, dV = P_1 \int_{-\infty}^{x_o} p_1(V)\, dV \,. \tag{3-90}$$

Receiver Structure

Define the discrete random variable

$$T(x) = \begin{cases} 1 \text{ if "1" sent, "0" decided} \\ -1 \text{ if "0" sent, "1" decided} \\ 0 \text{ if decision correct.} \end{cases} \tag{3-91}$$

Then

$$N(x) = E[T(x)]$$

$$N(x) = P_1 \int_{-\infty}^{x} p_1(V)\,dV - P_o \int_{x}^{\infty} p_o(V)\,dV \,. \tag{3-92}$$

The first integral is monotonically increasing and the second is monotonically decreasing with x so that N(x) is monotonically increasing and the requirement of Equation (2-8) is satisfied. Evaluating N(x) at x_o,

$$N(x_o) = P_1 \int_{-\infty}^{x_o} p_1(V)\,dV - P_o \int_{x_o}^{\infty} p_o(V)\,dV$$

$$N(x_o) = \beta - \alpha \,.$$

But, by the criterion,

$$\alpha = \beta$$

so that $N(x_o)$ is zero and Equation (2-9) is satisfied. Differentiating N(x) gives

$$N'(x) = P_1\,p_1(x) + P_o\,p_o(x) \tag{3-93}$$

so that Equation (2-10) is satisfied if $p_1(x_o)$ and $p_o(x_o)$ are not both zero. Equation (2-12) is satisfied since T(x) is finite for all x. Then, with $\{a_n\}$ properly defined, the recursive equation

$$x_{n+1} = x_n - a_n\, T(x) \qquad (3\text{-}94)$$

converges to the optimum threshold, defined by Equation (3-90), in mean square and with probability one as n becomes infinite. The adaptive threshold receiver is shown in Figure 18.

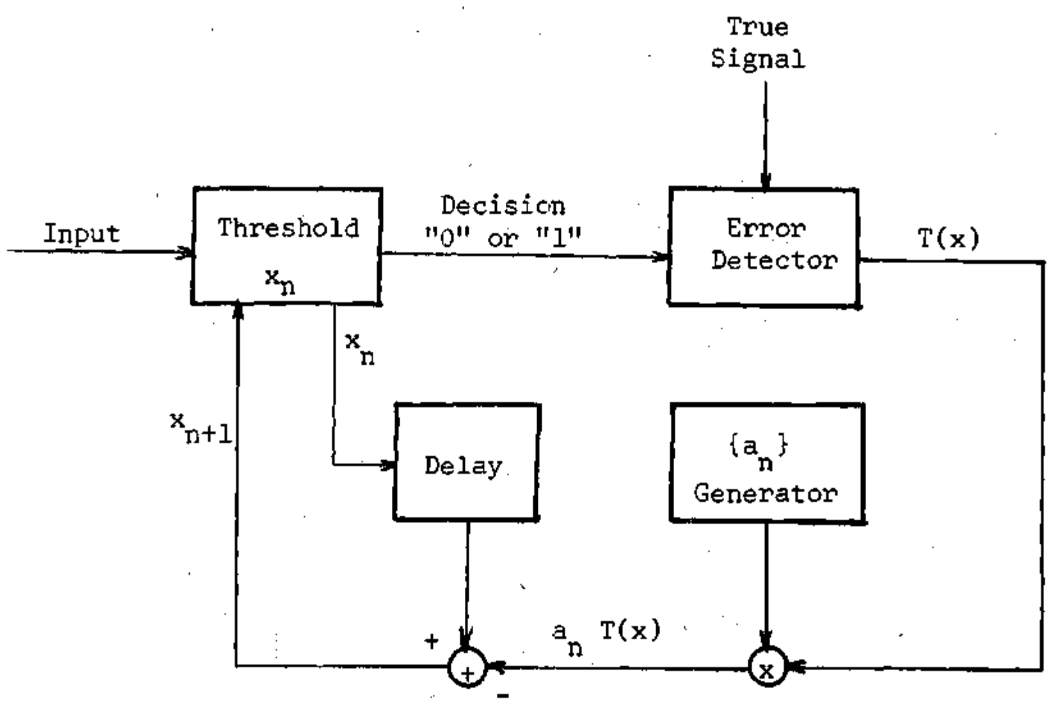


Figure 18. Adaptive Receiver for the Criterion:
$P_o\, P[1|0] = P_1\, P[0|1]$

For this criterion, T(x) may be considered an "error detector." If a "1" is transmitted and "0" decided, the threshold is moved down an amount a_n, thereby reducing the probability of this kind of error occurring. If a "0" is transmitted and "1" decided, the threshold is moved up an amount a_n and the probability of this kind of error is reduced. If either "1" or "0" is transmitted and a correct decision is made, the threshold is unchanged.

Convergence

Condition 1 of Sacks' theorem is satisfied by the definition of $N(x)$. Condition 2, 4(a) and 5 are satisfied because $N(x)$ and $T(x)$ are finite for all x. Condition 3 is satisfied by defining

$$\alpha_1 = P_1\, p_1(x_o) + P_o\, p_o(x_o)\ . \tag{3-95}$$

Using Condition 4(b) to determine γ^2,

$$T^2(x) = \begin{cases} 1 \text{ if decision incorrect} \\ 0 \text{ if decision correct} \end{cases}$$

so that

$$\gamma^2 = \lim_{x \to x_o} E[T^2(x)]$$

$$\gamma^2 = P_o \int_{x_o}^{\infty} p_o(V)\, dV + P_1 \int_{-\infty}^{x_o} p_1(V)\, dV\ .$$

However, from the criterion,

$$P_o \int_{x_o}^{\infty} p_o(V)\, dV = P_1 \int_{-\infty}^{x_o} p_1(V)\, dV ,$$

$$\alpha = \beta ,$$

so that

$$\gamma^2 = 2\alpha = 2\beta . \tag{3-96}$$

Then with

$$a_n = A/n ,$$

where

$$A > 1/2[P_o\, p_o(x_o) + P_1\, p_1(x_o)] ,$$

$$\mathrm{Var}\ (x_n - x_o) = \frac{1}{n}\ \frac{2\alpha A^2}{2\ A[P_o\, p_o(x_o) + P_1\, p_1(x_o)] - 1} . \tag{3-97}$$

If

$$A = 1/[P_o\, p_o(x_o) + P_1\, p_1(x_o)] ,$$

$$\mathrm{Min\ Var}\ (x_n - x_o) = \frac{1}{n} \frac{2\alpha}{[P_o\, p_o(x_o) + P_1\, p_1(x_o)]^2} . \tag{3-98}$$

For this criterion, unlike the earlier ones, there is no parameter, such as α or β, externally controlled at the receiver. The only controllable parameter is the amplitude of the transmitted signal. Therefore, the minimum error variance will be evaluated and plotted as a function of this parameter.

If the input is uniformly distributed when either the "0" or "1" symbol is transmitted, with

$$p_o(V) = \begin{cases} \frac{1}{2b} \text{ for } |V| \leq b \\ 0 \text{ for } |V| > b \end{cases} \tag{3-99}$$

and

$$p_1(v) = \begin{cases} \frac{1}{2b} \text{ for } |V - M_1| \leq b \\ 0 \text{ for } |V - M_1| > b \ , \end{cases} \tag{3-100}$$

then

$$E[V|\text{"0" Transmitted}] = 0 \ ,$$

$$E[V|\text{"1" Transmitted}] = M_1 \ ,$$

and the variance when either signal is present is $b^2/3$. Solving Equation (3-90), the optimum threshold as a function of M_1 is

$$x_o = b + P_1 \ (M_1 - 2b) \tag{3-101}$$

so that

$$\alpha = \frac{P_o P_1}{2} (2 - \frac{M_1}{b}) \tag{3-102}$$

and

$$\text{Min Var } (x_n - x_o) = \frac{b^2}{n} 4 P_o P_1 (2 - \frac{M_1}{b}) . \tag{3-103}$$

Thus, the normalized error variance varies from $8P_oP_1$, when M_1 is zero, to zero, when M_1 is 2b. For values of M_1 equal to or greater than 2b, the densities do not overlap and a range of x_o exists for which no decision errors are made.

If the input has a gaussian distribution when either the "0" or "1" symbol is transmitted, with $p_o(V)$ given by Equation (3-17) and

$$p_1(V) = \frac{1}{\sqrt{2\pi}\ \sigma} \exp \left[- \frac{(V - M_1)^2}{2\sigma^2} \right] \tag{3-104}$$

then

$$E[V|\text{"0" Transmitted}] = 0 ,$$

$$E[V|\text{"1" Transmitted}] = M_1 ,$$

and

$$\sigma_o^2 = \sigma_1^2 = \sigma^2 .$$

Defining

$$f_o(x_o) = \frac{1}{\sqrt{2\pi}} \exp \{- x_o^2/2\sigma^2\}$$

and

$$f_1(x_o) = \frac{1}{\sqrt{2\pi}} \exp \left[- \frac{(x_o - M_1)^2}{2\sigma^2}\right] ,$$

$$\text{Min Var } (x_n - x_o) = \frac{\sigma^2}{n} \frac{2\alpha}{[P_o f_o(x_o) + P_1 f_1(x_o)]^2} . \qquad (3\text{-}105)$$

Using standard tables of the normal distribution function, Equation (3-90) can be solved for x_o/σ, α determined and the minimum error variance expression evaluated as a function of M_1/σ. The result is shown in Figure 19. Due to the symmetry of the gaussian densities, the rate of convergence is symmetrical with respect to P_o; that is, the minimum error variance for

$$P_o = P_o'$$

is the same as for

$$P_o = 1 - P_o' .$$

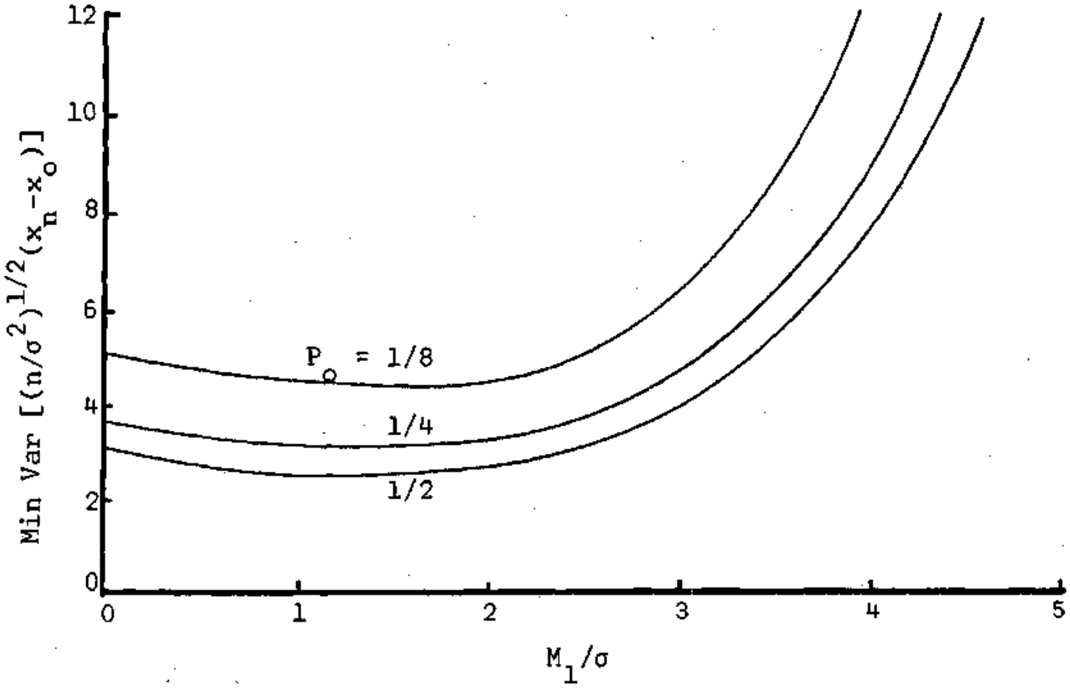


Figure 19. Minimum Error Variance for Gaussian Density and the Criterion: $P_o\ P[1|0] = P_1\ P[0|1]$

When the receiver employs a narrowband filter and envelope detector, such that $p_o(V)$ and $p_1(V)$ are given by Equations (3-25) and (3-73), respectively, the optimum threshold is the solution to

$$\left(\frac{P_o}{P_1}\right)\exp\left[-\frac{x_o^2}{2\sigma^2}\right] = 1 - Q\left(B, \frac{x_o}{\sigma}\right) . \tag{3-106}$$

Defining

$$u_o = x_o/\sigma , \tag{3-107}$$

$$B = A_1/\sigma , \tag{3-108}$$

$$f_o(u_o) = \frac{u_o}{\sigma} \exp\left[-\frac{u_o^2}{2}\right] , \qquad (3\text{-}109)$$

$$f_1(u_o) = \frac{u_o}{\sigma} \exp\left[-\frac{u_o^2 + B^2}{2}\right] I_o(Bu_o) , \qquad (3\text{-}110)$$

then

$$\alpha = P_o \exp\left[-\frac{u_o^2}{2}\right] , \qquad (3\text{-}111)$$

$$\beta = P_1[1 - Q(B,u_o)] , \qquad (3\text{-}112)$$

and

$$\text{Min Var } (x_n - x_o) = \frac{\sigma^2}{n} \; \frac{2\alpha}{[P_o \; f_o(u_o) + P_1 \; f_1(u_o)]^2} . \qquad (3\text{-}113)$$

The normalized optimum threshold is determined from Equation (3-106), using Marcum's (18) tabulation, and the minimum error variance evaluated as a function of A_1/σ. The result is shown in Figure 20.

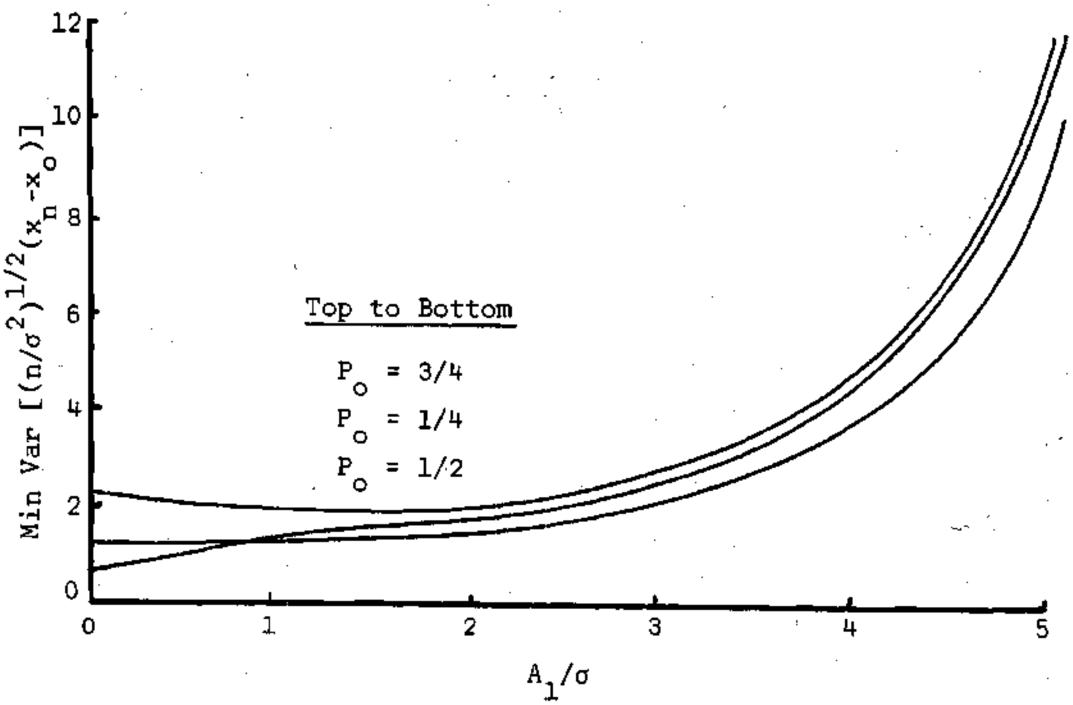


Figure 20. Minimum Error Variance for Ricean Density and the Criterion: $P_o\ P[1|0] = P_1\ P[0|1]$

Criterion: $P[1|0] = P[0|1]$

Instead of requiring that the average probabilities of the two kinds of error are equal, performance may be considered optimum when the conditional probabilities are equal. The optimum threshold is then defined by the equation

$$\int_{x_o}^{\infty} p_o(V)\ dV = \int_{-\infty}^{x_o} p_1(V)\ dV\ . \qquad (3\text{-}114)$$

Receiver Structure

As for other criteria not depending on the a priori probabilities, two methods of generating T(x) are available, with resulting different receiver structures and convergence rates.

$\underline{P_o \text{ Known}}$. Define the discrete random variable

$$T(x) = \begin{cases} \frac{1}{P_1} \text{ if "1" sent, "0" decided} \\ -\frac{1}{P_o} \text{ if "0" sent, "1" decided} \\ 0 \text{ if decision correct} \end{cases}$$

so that

$$E[T(x)] = \frac{1}{P_1} [P_1 \int_{-\infty}^{x} p_1(V)\, dV] - \frac{1}{P_o} [P_o \int_{x}^{\infty} p_o(V)\, dV] ,$$

and

$$N(x) = \int_{-\infty}^{x} p_1(V)\, dV - \int_{x}^{\infty} p_o(V)\, dV . \qquad (3\text{-}115)$$

Differentiating $N(x)$,

$$N'(x) = p_1(x) + p_o(x) \qquad (3\text{-}116)$$

so that, if $p_1(x_o)$ and $p_o(x_o)$ are not both zero and $\{a_n\}$ is properly chosen, all of the requirements are satisfied. The threshold setting is defined by

$$x_{n+1} = x_n - a_n\, T(x) , \qquad (3\text{-}117)$$

which converges to x_o as n approaches infinity. The receiver is similar to the one of Figure 18, except that the a priori probabilities are required as inputs in order to generate T(x).

P_o Unknown. Earlier methods of generating T(x), such that the solution of $N(x_o)$ is independent of the a priori probabilities, cannot be used here because both $p_o(V)$ and $p_1(V)$ are included in the criterion. Rather, for this case, T(x) will depend not only on the symbol and decision of the present interval, but also on the symbol of the previous interval. Define

$$T(x) = \begin{cases} 1 \text{ if "1" sent, "0" decided and} \\ \qquad \text{"0" sent during previous interval.} \\ -1 \text{ if "0" sent, "1" decided and} \\ \qquad \text{"1" sent during previous interval.} \\ 0 \text{ if otherwise.} \end{cases} \tag{3-118}$$

Then

$$N(x) = P_o[P_1 \int_{-\infty}^{x} p_1(V)\, dV] - P_1[P_o \int_{x}^{\infty} p_o(V)\, dV]$$

or

$$N(x) = P_o\, P_1[\int_{-\infty}^{x} p_1(V)\, dV - \int_{x}^{\infty} p_o(V)\, dV] \tag{3-119}$$

and

$$N'(x) = P_o\, P_1[p_1(x) + p_o(x)]\ . \tag{3-120}$$

If $p_o(x_o)$ and $p_1(x_o)$ are not both zero and $\{a_n\}$ is properly chosen, all of the requirements are satisfied and

$$x_{n+1} = x_n - a_n\, T(x) \tag{3-121}$$

converges to the optimum threshold as n becomes infinite. The receiver for this criterion is the same as the one of Figure 18 except for the definition of T(x). This definition allows the threshold to be adjusted only when a decision error is made *and* the transmitted symbol changed from the previous interval. Thus, the receiver trains itself on an alternating sequence of "0"s and "1"s, regardless of the actual transmitted sequence; thereby removing the a priori probabilities from the solution.

Convergence

P_o Known. With

$$\alpha_1 = p_1(x_o) + p_o(x_o) \tag{3-122}$$

all of the conditions of Sacks' theorem are satisfied. Then

$$T^2(x) = \begin{cases} \dfrac{1}{P_1^{\,2}} \text{ if "1" sent, "0" decided} \\ \dfrac{1}{P_o^{\,2}} \text{ if "0" sent, "1" decided} \\ 0 \quad \text{if otherwise.} \end{cases}$$

so that

$$\gamma^2 = \lim_{x\to x_o}\left[\frac{1}{P_1{}^2}\,[P_1 \int_{-\infty}^{x} p_1(V)\ dV] + \frac{1}{P_o{}^2}\,[P_o \int_{x}^{\infty} p_o(V)\ dV]\right]$$

$$\gamma^2 = \frac{1}{P_1}\int_{-\infty}^{x_o} p_1(V)\ dV + \frac{1}{P_o}\int_{x_o}^{\infty} p_o(V)\ dV\ .$$

But by the criterion,

$$\int_{-\infty}^{x_o} p_1(V)\ dV = \int_{x_o}^{\infty} p_o(V)\ dV\ ,$$

so that

$$\gamma^2 = \left(\frac{1}{P_1} + \frac{1}{P_o}\right)\int_{x_o}^{\infty} p_o(V)\ dV$$

$$\gamma^2 = \frac{1}{P_o\ P_1}\int_{x_o}^{\infty} p_o(V)\ dV$$

$$\gamma^2 = \frac{\alpha'}{P_o P_1}\ . \qquad (3\text{-}123)$$

With

$$a_n = A/n\ ,$$

where

$$A > 1/2[p_o(x_o) + p_1(x_o)] \ ,$$

$$\text{Var}\ (x_n - x_o) = \frac{1}{n}\ \frac{A^2\alpha'}{P_o\ P_1\ \{2A[p_o(x_o) + p_1(x_o)] - 1\}}\ , \quad (3\text{-}124)$$

and with

$$A = \frac{1}{[p_o(x_o) + p_1(x_o)]}\ ,$$

$$\text{Min Var}\ (x_n - x_o) = \frac{1}{n}\ \frac{\alpha'}{P_o P_1 [p_o(x_o) + p_1(x_o)]^2}\ . \qquad (3\text{-}125)$$

This is similar to the result, Equation (3-98), for the criterion

$$P_o\ P[1|0] = P_1\ P\ [0|1]\ .$$

In fact, if

$$P_o = P_1 = 1/2$$

for that criterion, the optimum threshold determined from Equation (3-90) is the same as determined from Equation (3-114) and Equation (3-98) becomes

$$\text{Min Var } (x_n - x_o) = \frac{1}{n} \frac{4\alpha'}{[p_o(x_o) + p_1(x_o)]^2} . \qquad (3\text{-}126)$$

The convergence properties for the present criterion can therefore be obtained from the results of the earlier criterion by scaling the amplitude of the P_o equal to one-half curves of Figures 19 and 20 by $\frac{1}{4P_oP_1}$, using the a priori values of interest.

<u>P_o Unknown</u>. With

$$\alpha_1 = P_oP_1[p_o(x_o) + p_1(x_o)] \qquad (3\text{-}127)$$

all of the conditions are satisfied and

$$T^2(x) = \begin{cases} 1 & \text{if decision incorrect and symbol changed} \\ 0 & \text{if otherwise} \end{cases}$$

so that

$$\gamma^2 = \lim_{x \to x_o} \left[P_o[P_1 \int_{-\infty}^{x} p_1(V)\, dV] + P_1[P_o \int_{x}^{\infty} p_o(V)\, dV] \right]$$

$$\gamma^2 = P_oP_1[\int_{-\infty}^{x_o} p_1(V)\, dV + \int_{x_o}^{\infty} p_o(V)\, dV]$$

$$\gamma^2 = 2P_oP_1\alpha' . \qquad (3\text{-}128)$$

With

$$a_n = A/n \ ,$$

where

$$A > 1/2P_oP_1[p_o(x_o) + p_1(x_o)] \ ,$$

$$\text{Var}\ (x_n - x_o) = \frac{1}{n}\ \frac{2\ A^2\ P_o\ P_1\ \alpha'}{2AP_oP_1[p_o(x_o) + p_1(x_o)] - 1}\ , \qquad (3\text{-}129)$$

or with

$$A = 1/P_oP_1[p_o(x_o) + p_1(x_o)] \ ,$$

$$\text{Min Var}\ (x_n - x_o) = \frac{1}{n}\ \frac{2\alpha'}{P_oP_1[p_o(x_o) + p_1(x_o)]^2}\ . \qquad (3\text{-}130)$$

This is just twice the result for the case where P_o is known, Equation (3-125). The difference in the two systems is that adjustment of the former occurs every time a decision error is made, whereas the latter is adjusted only when a decision error is made *and* the transmitted symbol changed. By adjusting less often, the required training time is doubled.

<u>Criterion: Minimize $\{P_o\ P[1|0] + P_1\ P[0|1]\}$</u>

The criterion most frequently used in communication applications requires that the average probability of error be minimized. The average probability of error is

$$R(x) = \alpha(x) + \beta(x)$$

$$R(x) = P_o \int_x^{\infty} p_o(V)\ dV + P_1 \int_{-\infty}^{x} p_1(V)\ dV \tag{3-131}$$

so that the optimum threshold is defined by

$$\min_x R(x) = R(x_o) = P_o \int_{x_o}^{\infty} p_o(V)\ dV + P_1 \int_{-\infty}^{x_o} p_1(V)\ dV\ . \tag{3-132}$$

Up to this point, all of the criteria considered were such that the problem could be formulated as one of finding the point at which a function is zero. Application of the Robbins-Monro method to such problems is relatively straightforward. This criterion requires the minimization of a function and the minimum value is greater than zero for applications of interest. However, a necessary condition that $R(x_o)$ be a minimum is

$$R'(x_o) = 0 \tag{3-133}$$

so that the optimum threshold is the solution to

$$P_o\, p_o(x_o) = P_1\, p_1(x_o) \ , \tag{3-134}$$

if $R'(x_o)$ is defined. The Robbins-Monro method can then be applied by formulating the problem as an approximation to the derivative of $R(x)$ and finding the zero point of that function.

Another approach is provided by the stochastic approximation theorem of Kiefer and Wolfowitz (19). Their method is directly concerned with maximizing or minimizing a function such as $R(x)$. Both methods require that $R(x)$ has a unique minimum just as the earlier functions were required to have unique zeros.

Robbins-Monro Method

Receiver Structure. Define the discrete random variable

$$T(x) = \frac{1}{2c}\,[y(x + c) - y(x - c)]$$

where

$$y(x) = \begin{cases} 1 & \text{if decision incorrect} \\ 0 & \text{if decision correct,} \end{cases}$$

and c is an arbitrary constant. $y(x + c)$ is the 0,1 random variable generated when the input signal is compared to a threshold set at $x + c$. $y(x - c)$ is the 0,1 random variable generated when the input signal is compared to a threshold set at $x - c$. Then

$$N(x) = E[T(x)]$$

$$N(x) = \frac{1}{2c}\left[\left[P_o \int_{x+c}^{\infty} p_o(V)\, dV + P_1 \int_{-\infty}^{x+c} p_1(V)\, dV\right.\right. \qquad (3\text{-}135)$$

$$\left.\left. - \left[P_o \int_{x-c}^{\infty} p_o(V)\, dV + P_1 \int_{-\infty}^{x-c} p_1(V)\, dV\right]\right]\right.$$

$$N(x) = \frac{R(x + c) - R(x - c)}{2\ c}\ , \qquad (3\text{-}136)$$

which for small c becomes

$$N(x) \cong R'(x) = P_1\, p_1(x) - P_o\, p_o(x)\ . \qquad (3\text{-}137)$$

Since $N(x)$ is only an approximation to $R'(x)$, the value of x that makes $N(x)$ zero will not be x_o but some slightly different value, x_o'. Only if $R(x)$ is symmetrical near x_o will x_o' be equal to x_o. By decreasing the value of c, the approximation can be made better and a more accurate estimate of x_o obtained. However, it will be shown later that this also increases the required training time.

Differentiating Equation (3-136),

$$N'(x) = \frac{R'(x + c) - R'(x - c)}{2\ c} \qquad (3\text{-}138)$$

which for small c becomes

$$N'(x) \cong R''(x) = P_1\, p_1'(x) - P_o\, p_o'(x) \ . \tag{3-139}$$

The requirements on $N(x)$, stated in Equations (2-8), (2-9) and (2-10), are satisfied if

$$R''(x) \geq 0 \quad \text{for } x \neq x_o \ , \tag{3-140}$$

$$R'(x_o) = 0 \ , \tag{3-141}$$

and

$$R''(x_o) > 0 \ . \tag{3-142}$$

The last two requirements on $R(x)$ are satisfied if $R(x)$ has a unique minimum at x_o and $R'(x_o)$ is defined. When the input densities are uniform,

$$p_o'(x_o) = p_1'(x_o) = 0$$

so that Equation (3-142) is not satisfied. In fact, for the uniform input density functions, $R(x)$ may only take one of the three forms shown in Figure 21. For the top form, a unique value of x_o does not exist. For the other two forms, $R'(x_o)$ is not defined. Therefore, a receiver based on the Robbins-Monro theorem cannot be used to minimize the average probability of error when the input signals are uniformly distributed. This is not a severe limitation; however, since it is

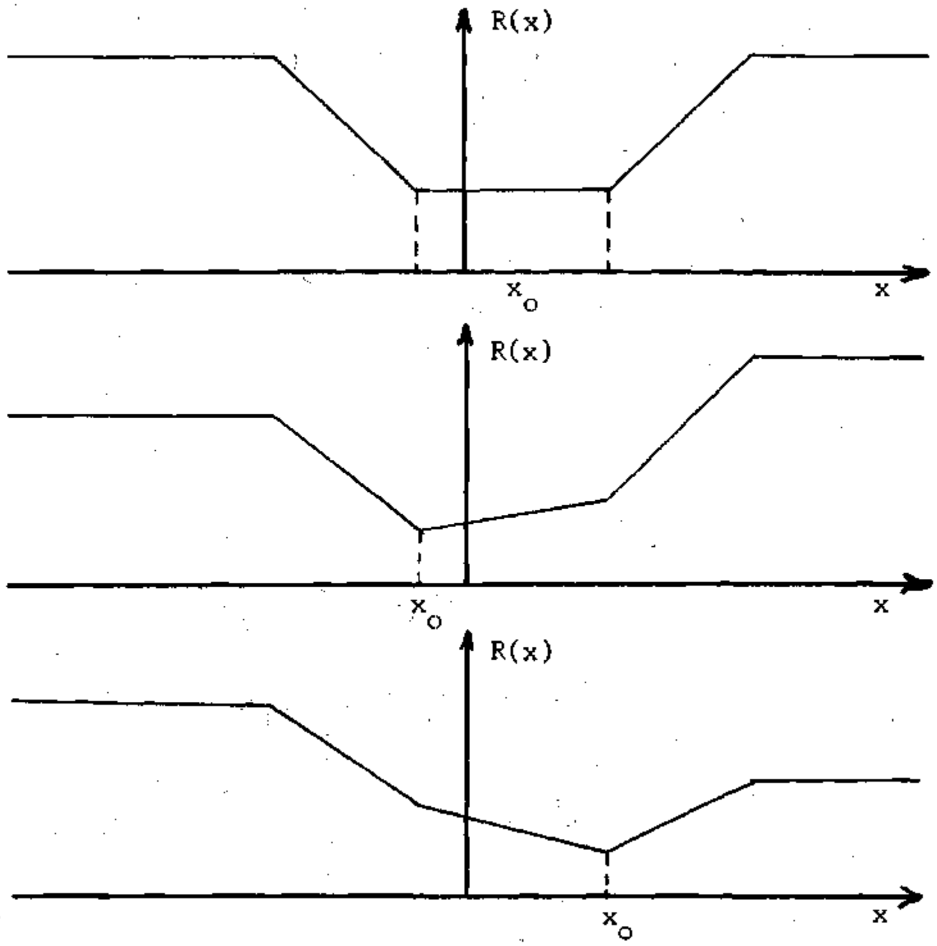


Figure 21. R(x) for Uniformly Distributed Input Signals

unlikely that a perfectly flat amplitude distribution could ever arise in realistic applications.

The requirement of Equation (3-140) is more significant. A R(x), and its first two derivatives, typically encountered in practice is shown in Figure 22. R(x) asymptotically approaches a finite value as x approaches plus and minus infinity. Therefore, R'(x) must asymptotically approach zero as x approaches plus and minus infinity. Since R'(x) must be zero at only one point, x_o, it must be negative for

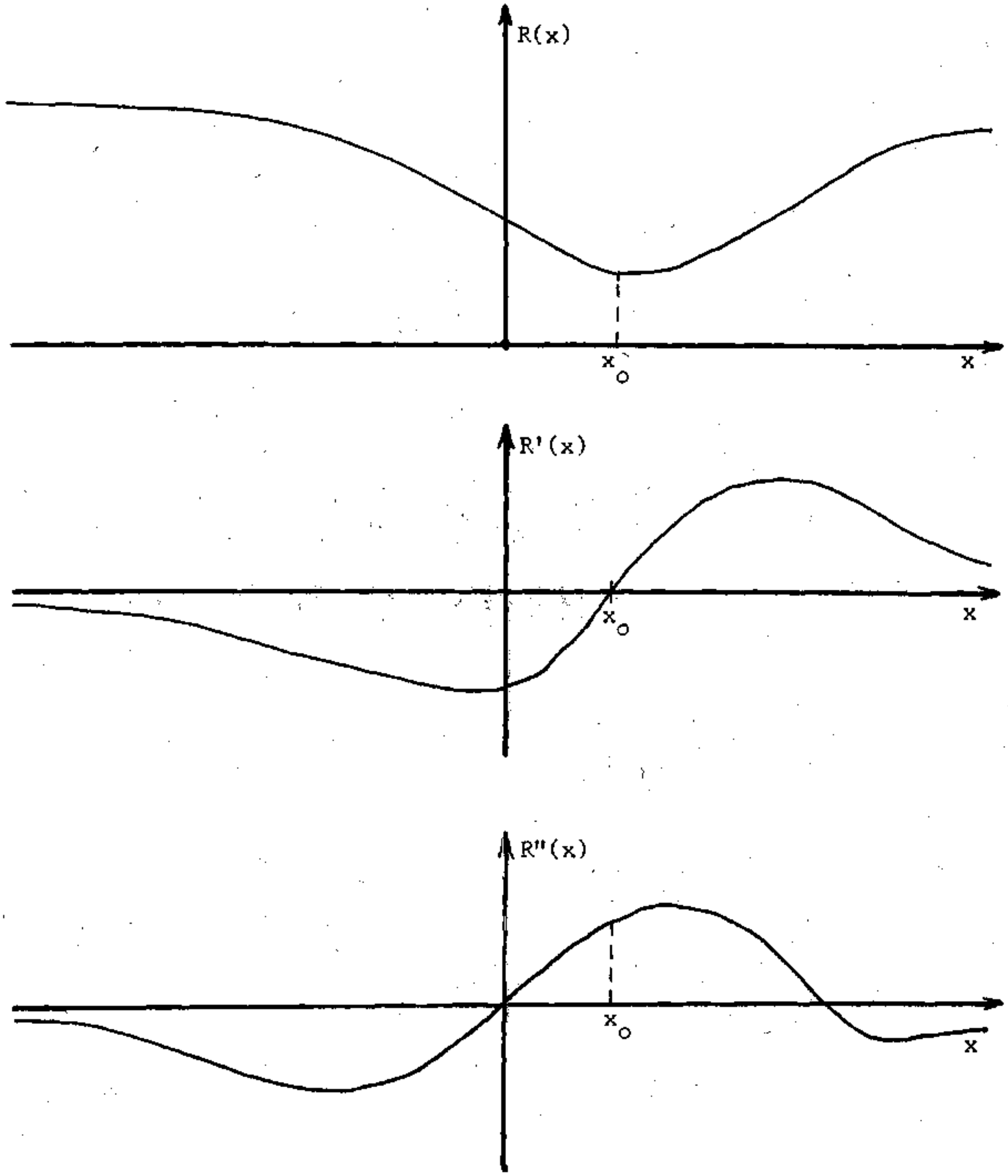


Figure 22. A Typical R(x) and Its First Two Derivatives

x less than x_o and positive for x greater than x_o. $R'(x)$ must then have a minimum for some x less than x_o and a maximum for some x greater than x_o. Hence, its slope, $R''(x)$, must be zero at one point above and one point below x_o. The result is two regions in which $R''(x)$ is negative and does not satisfy Equation (3-140). Normally these regions will be relatively far removed from x_o. In order to assure that the receiver converges to x_o', some preliminary system investigation must be made to determine a region in which x_o would be expected to lie. The limits of this region would then be used to constrain the threshold setting during training to the proper region.

These problems can be reduced somewhat if Robbins and Monro's first theorem, stated in Appendix I, is used in place of their second theorem, stated in Chapter II. The requirements of Equations (3-140), (3-141) and (3-142) are then replaced by the requirement that $R'(x)$ be non-zero for finite values of x, other than x_o. For input density functions which extend to plus and minus infinity, such as the gaussian, $R'(x)$ asymptotically approaches zero as x approaches plus and minus infinity, as shown in Figure 22. The less restrictive requirement is satisfied and no special considerations are necessary at the receiver.

For input density functions which are zero over some range, such as the uniform and Ricean, $R(x)$ will be constant over that range and $R'(x)$ will be zero; even the less restrictive requirements are not satisfied. However, if the receiver includes an envelope detector, it will be known beforehand that the optimum threshold must be positive. During training the receiver must include a constraint to allow only those threshold adjustments which result in positive threshold settings.

The receiver will then be assured of converging to x_o'.

When the input is uniformly distributed such that there is not a unique x_o, neither theorem is satisfied and a converging receiver cannot be obtained. However, the less restrictive theorem does not require that $N(x_o)$, thus $R'(x_o)$, be defined, so that the receiver will converge if $R(x)$ is of either of the lower two types of Figure 21 and a constraint is used to keep the threshold settings in the proper region.

The requirement of Equation (2-12) is satisfied because $T(x)$ takes on only finite values. Then with $\{a_n\}$ properly defined, the iterative process,

$$x_{n+1} = x_n - \frac{a_n}{2c}\left[y(x + c) - y(x - c)\right] \tag{3-143}$$

converges to x_o' as n becomes infinite. The receiver block diagram is shown in Figure 23.

The receiver may operate with both $y(x + c)$ and $y(x - c)$ generated from the same input observation or they may be generated alternately, using independent observations. Using the first method, decisions are made by both thresholds during each signal interval, compared to the true signal, $y(x + c)$ and $y(x - c)$ generated and combined to form $T(x)$, and both thresholds adjusted for the next signal interval. Using the second method, the input is applied to only one of the thresholds, resulting in a corresponding $y(x)$.

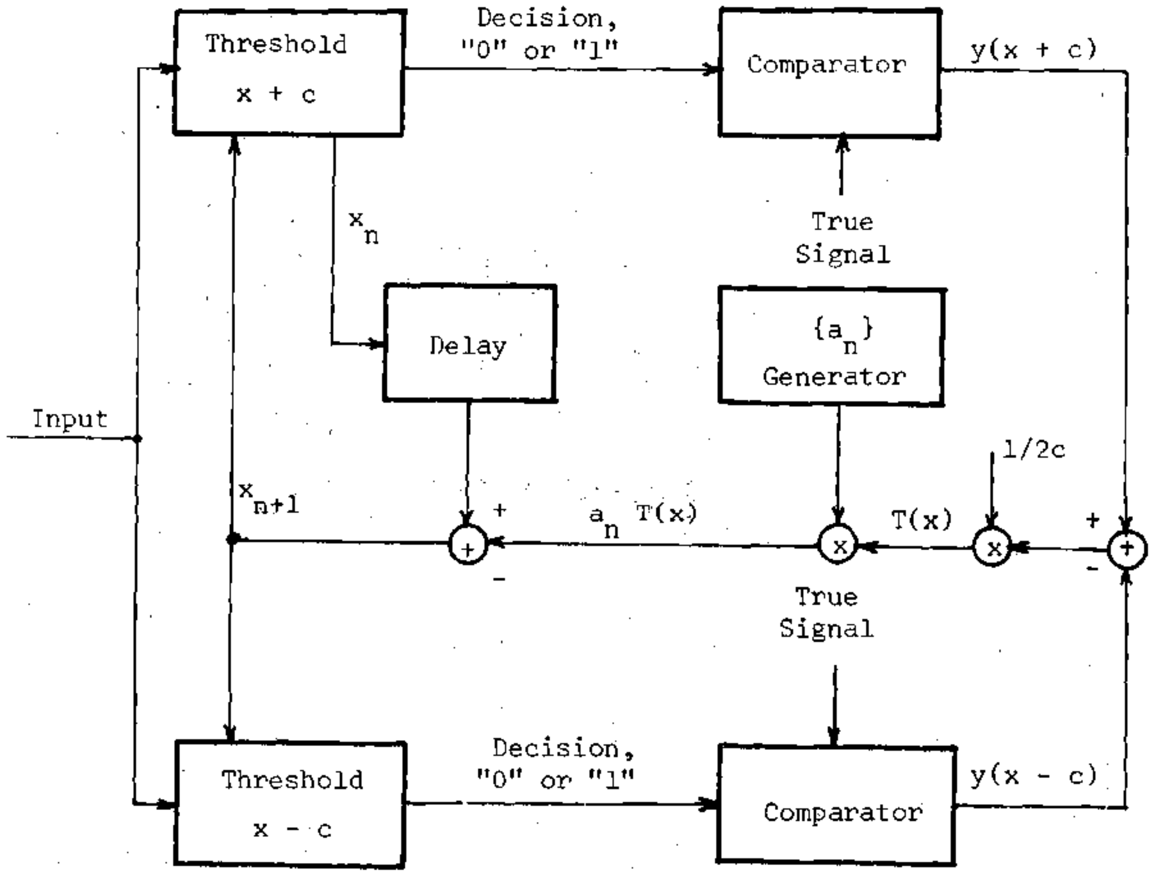


Figure 23. Adaptive Receiver for the Criterion: Minimize $\{P_o\ P[1|0] + P_1\ P[0|1]\}$

During the next signal interval, the input is applied to the other threshold, resulting in a generation of the other y(x). Then y(x + c) and y(x - c) are combined to form T(x) and both thresholds adjusted for the next two signal intervals. For the second case, n, the number of iterations is one-half the number of observations. However, a comparison of the resulting error variances is required before selecting the training method.

<u>Convergence When One Observation Used.</u> Taking into account any necessary receiver constraints, Condition 1 of Sacks' theorem is satisfied by formulating the problem in accordance with either Equations (2-8), (2-9), and (2-10) or Equations (2-8), (A1-1), and (A1-2). Conditions 2, 4(a) and 5 are satisfied since T(x) and N(x) are finite for all x. Condition 3 is satisfied if c is small and

$$\alpha_1 = P_1 \, p_1'(x_o') - P_o \, p_o'(x_o') \, . \tag{3-144}$$

Proceeding to find γ^2,

$$T^2(x) = \frac{1}{4c^2} [y^2(x + c) - 2y(x + c) \, y(x - c) + y^2(x - c)],$$

but

$$y^2(x + c) = y(x + c) \ ,$$

$$y^2(x - c) = y(x - c) \ ,$$

so that

$$T^2(x) = \frac{1}{4c^2} [y(x + c) - 2y(x + c) \, y(x - c) + y(x - c)]$$

and

$$\gamma^2 = \lim_{x \to x_o'} E\left[\left(\frac{1}{4c^2}\right)[y(x + c) - 2y(x + c)\ y(x - c) + y(x - c)]\right] . \quad (3\text{-}145)$$

Taking the expectation of the first term,

$$E[y(x + c)] = P_o \int_{x+c}^{\infty} p_o(V)\ dV + P_1 \int_{-\infty}^{x+c} p_1(V)\ dV$$

so that

$$E[y(x + c)] = R(x + c)$$

and

$$\lim_{x \to x_o'} E[y(x + c)] = R(x_o' + c) . \quad (3\text{-}146)$$

Similarly, the last term becomes

$$E[y(x - c)] = R(x - c)$$

$$\lim_{x \to x_o'} E[y(x - c)] = R(x_o' - c) . \quad (3\text{-}147)$$

At this point the distinction between the two methods of generating $y(x + c)$ and $y(x - c)$ must be noted. The expectation of the product term will depend on the method used. From the definitions of $y(x + c)$ and $y(x - c)$ and when the same observation is used to generate both

$$y(x + c)\, y(x - c) = \begin{cases} 1 \text{ if "1" sent and received signal below } x - c \\ 1 \text{ if "0" sent and received signal above } x + c \\ 0 \text{ if otherwise.} \end{cases} \tag{3-148}$$

so that

$$E[2y(x + c)\, y(x - c)] = 2[P_1 \int_{-\infty}^{x-c} p_1(V)\, dV + P_o \int_{x+c}^{\infty} p_o(V)\, dV]$$

or

$$E[2y(x + c)\, y(x - c)] = 2[\beta(x - c) + \alpha(x + c)]$$

and

$$\lim_{x \to x_o'} E[2y(x + c)\, y(x - c)] = 2[\alpha(x_o' + c) + \beta(x_o' - c)] \;. \tag{3-149}$$

Noting that the problem was formulated as

$$N(x) = \frac{R(x + c) - R(x - c)}{2\, c}$$

so that, evaluating at x_o',

$$R(x_o' + c) = R(x_o' - c)\;, \tag{3-150}$$

γ^2 becomes

$$\gamma^2 = \frac{1}{4c^2}\{2R(x_o' + c) - 2[\alpha(x_o' + c) + \beta(x_o' - c)]\}$$

$$\gamma^2 = \frac{1}{2c^2}[\alpha(x_o' + c) + \beta(x_o' + c) - \alpha(x_o' + c) - \beta(x_o' - c)]$$

$$\gamma^2 = \frac{1}{c}\left[\frac{\beta(x_o' + c) - \beta(x_o' - c)}{2\ c}\right]$$

$$\gamma^2 \cong \frac{1}{c}\left[\frac{d}{dx}\beta\ (x)\Big|_{x=x_o'}\right]$$

$$\gamma^2 \cong \frac{1}{c} P_1\ p_1(x_o') \cong \frac{1}{c} P_o\ p_o(x_o')\ . \qquad (3\text{-}151)$$

Then, with

$$a_n = A/n\ ,$$

where

$$A > 1/2[P_1\ p_1'(x_o') - P_o\ p_o'(x_o')]\ ,$$

$$\mathrm{Var}\ (x_n - x_o') \cong \frac{1}{nc}\ \frac{A^2\ P_o\ p_o(x_o')}{2A[P_1\ p_1'(x_o') - P_o\ p_o'(x_o')] - 1}, \qquad (3\text{-}152)$$

or if

$$A = \frac{1}{[P_1 \, p_1'(x_o') - P_o \, p_o'(x_o')]} ,$$

$$\text{Min Var } (x_n - x_o') \cong \frac{1}{nc} \; \frac{P_o \, p_o(x_o')}{[P_1 p_1'(x_o') - P_o \, p_o'(x_o')]^2} . \quad (3\text{-}153)$$

When the input density functions are uniform, $N'(x_o)$ is zero so that a finite value of A, satisfying the condition

$$A > 1/2N'(x_o) ,$$

does not exist. Therefore, even when R(x) is of the type which has a unique minimum, the convergence properties cannot be determined from Sacks' theorem. In this case, the receiver converges but its performance cannot be evaluated. This situation will occur whenever the criterion of interest involves minimization and the input signals are uniformly distributed.

Convergence When Two Observations Used. The procedure is the same as for when one observation is used, up to Equation (3-148). When separate observations are used, and the observations are independent due to a basic assumption discussed in Chapter II.

$$E[2y(x + c) \, y(x - c)] = 2E[y(x + c)] \, E[y(x - c)]$$

$$E[2y(x + c) \, y(x - c)] = 2 \, R(x + c) \, R(x - c)$$

and

$$\lim_{x \to x_o'} 2E[y(x + c)\, y(x - c)] = 2\, R(x_o' + c)\, R(x_o' - c) \; .$$

Using Equation (3-150), this becomes

$$\lim_{x \to x_o'} 2E[y(x + c)\, y(x - c)] = 2\, R^2(x_o' + c)$$

so that

$$\gamma^2 = \frac{1}{4c^2} [2R(x_o' + c) - 2R^2(x_o' + c)]$$

$$\gamma^2 = \frac{1}{2c^2} [R(x_o' + c) - R^2(x_o' + c)] \; . \qquad (3\text{-}154)$$

Then for small c and

$$a_n = A/n \; ,$$

where

$$A > 1/2[P_1\, p_1'(x_o') - P_o\, p_o'(x_o')] \; ,$$

$$\mathrm{Var}\,(x_n - x_o') \cong \frac{1}{2nc^2}\; \frac{A^2[R(x_o') - R^2(x_o')]}{2A[P_1\, p_1'(x_o') - P_o\, p_o'(x_o')] - 1} \; , \qquad (3\text{-}155)$$

or, if

$$A = 1/[P_1\, p_1'(x_o') - P_o\, p_o'(x_o')] \;,$$

$$\text{Min Var }(x_n - x_o') \cong \frac{1}{nc^2}\;\frac{R(x_o') - R^2(x_o')}{2[P_1\, p_1'(x_o') - P_o\, p_o'(x_o')]^2} \;. \quad (3\text{-}156)$$

For both methods the required training time increases as c is decreased. A trade-off is therefore required between the "$x_o - x_o'$" error and the "$x_n - x_o'$" error for a given training time. The primary difference between error variances for the two methods is the extra factor of c in the denominator of Equation (3-156). This factor can significantly increase the training time when an accurate estimate of x_o is required.

When the input signals have the gaussian densities given by Equations (3-17) and (3-104),

$$p_o'(x_o') = \frac{1}{\sqrt{2\pi}\;\sigma}\left(-\frac{x_o}{\sigma^2}\right) \exp\{-\,x_o'^2/2\sigma^2\} \;, \quad (3\text{-}157)$$

and

$$p_1'(x_o') = \frac{1}{\sqrt{2\pi}\;\sigma}\left[-\frac{(x_o' - M_1)}{\sigma^2}\right] \exp\left[-\frac{(x_o' - M_1)^2}{2\sigma^2}\right] . \quad (3\text{-}158)$$

Defining

$$u_o' = x_o'/\sigma \;, \quad (3\text{-}159)$$

$$f_o(u_o') = \sigma p_o(u_o') \ , \tag{3-160}$$

$$f_o'(u_o') = \sigma^2\ p_o'(u_o') \ , \tag{3-161}$$

and

$$f_1'(u_o') = \sigma^2\ p_o'(u_o') \ , \tag{3-162}$$

the minimum error variance when one observation is used becomes

$$\text{Min Var}\ (x_n - x_o') \cong \frac{\sigma^3}{nc}\ \frac{P_o\ f_o(u_o')}{[P_1\ f_1'(u_o') - P_o\ f_o'(u_o')]^2} \tag{3-163}$$

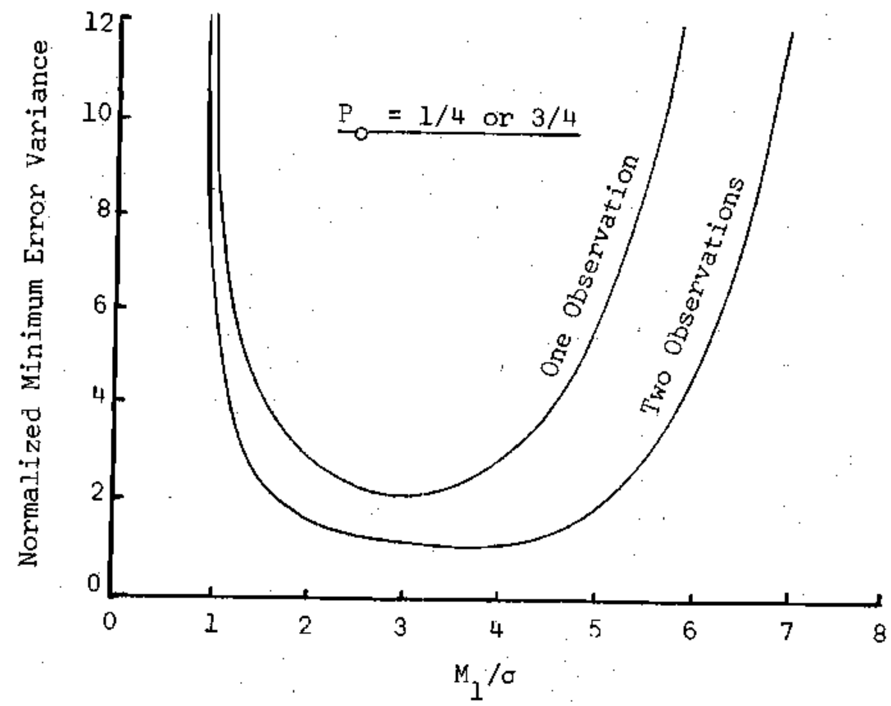
and the minimum error variance when two observations are used becomes

$$\text{Min Var}\ (x_n - x_o') \cong \frac{\sigma^4}{nc^2}\ \frac{R(u_o') - R^2(u_o')}{2[P_1\ f_1'(u_o') - P_o\ f_o'(u_o')]^2} \ . \tag{3-164}$$

The normalized optimum threshold is obtained from Equation (3-134) by taking the natural logarithm and rearranging terms, with the result

$$u_o = \frac{M_1}{2\sigma} + \left(\frac{M_1}{\sigma}\right)^{-1}\ \ln \frac{P_o}{P_1} \ . \tag{3-165}$$

Equations (3-163) and (3-164) are normalized by σ^3/nc and σ^4/nc^2, respectively, and plotted in Figure 24 as a function of M_1/σ. Both



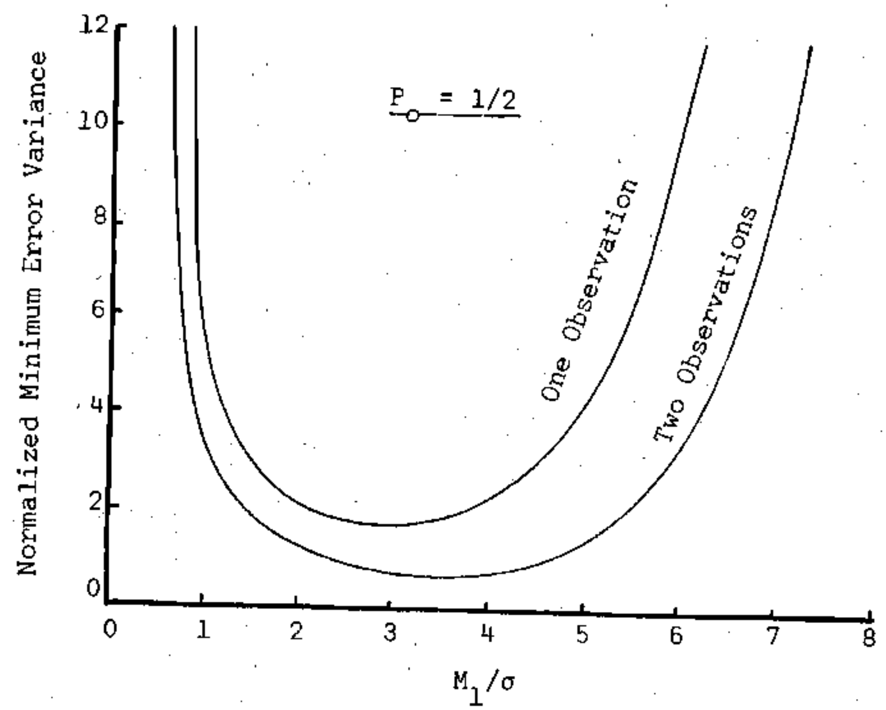


Figure 24. Minimum Error Variance for Gaussian Density and the Criterion: Minimize $\{P_o\, P[1|0] + P_1\, P[0|1]\}$

expressions increase without bound as M_1/σ becomes large; however, when two observations are used the increase is at a slower rate. The asymptotic approximations to the two expressions as M_1/σ becomes large are derived and compared in Appendix II.

When the input signals have the Ricean densities given by Equations (3-25) and (3-73), and using the definitions of Equations (3-107) and (3-108),

$$p_o'(u_o') = \frac{1}{\sigma^2} \exp\{-u_o'^2/2\} [1 - u_o'^2] \tag{3-166}$$

$$p_1'(u_o') = \frac{1}{\sigma^2} \exp\left[-\frac{(u_o'^2 + B^2)}{2}\right]\left[[1 - u_o'^2]\, I_o(Bu_o') + Bu_o'\, I_1(Bu_o')\right] \tag{3-167}$$

where $I_1(x)$ is the modified Bessel function of the first kind of order one and is related to $I_o(x)$ by (20)

$$I_o'(x) = I_1(x) \; . \tag{3-168}$$

The minimum error variance when one observation is used becomes

$$\text{Min Var } (x_n - x_o') \cong \frac{\sigma^3}{nc} \frac{[I_o(Bu_o')/I_1(Bu_o')]^2}{P_o\, u_o'\, B^2 \exp\{-u_o'^2/2\}} \tag{3-169}$$

and when two observations are used, becomes

$$\text{Min Var } (x_n - x_o') \cong \frac{\sigma^4}{nc^2} \frac{[R(u_o') - R^2(u_o')]\,[I_o(Bu_o')/I_1(Bu_o')]^2}{P_o^{\,2}\,(Bu_o')^2 \exp\{-u_o'^2\}} . \quad (3\text{-}170)$$

The corresponding normalized expressions are evaluated as functions of A_1/σ and plotted in Figure 25. Noting that the Ricean densities are similar to the gaussian densities for large A_1/σ and that $I_o(x)/I_1(x)$ approaches one as x becomes infinite, arguments similar to those of Appendix II can be used to obtain large A_1/σ approximations. The results, when one observation is used, is

$$\text{Min Var } (x_n - x_o') \cong \frac{\sigma^3}{nc} \frac{1}{P_o\, u_o'\, B^2 \exp\{-u_o'^2/2\}} \quad (3\text{-}171)$$

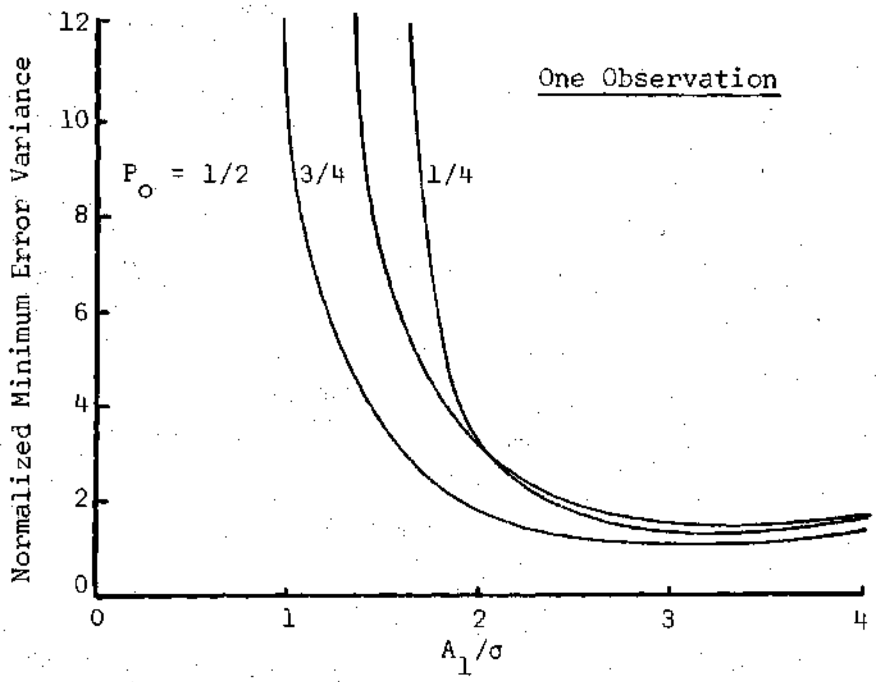
and, when two observations are used, is

$$\text{Min Var } (x_n - x_o') \cong \frac{\sigma^4}{nc^2} \frac{1}{P_o\, (u_o'B)^2 \exp\{-u_o'^2/2\}} . \quad (3\text{-}172)$$

Again, the difference is an extra factor of u_o' in the denominator of the latter, reducing its rate of increase with A_1.

Kiefer-Wolfowitz Method

The Robbins-Monro theorem proved not to be suited for direct application to the minimum probability of error criterion. In addition, the resulting receiver converged to a threshold setting slightly different from the optimum setting. Kiefer and Wolfowitz (19) later proved a stochastic approximation theorem for finding the point at which



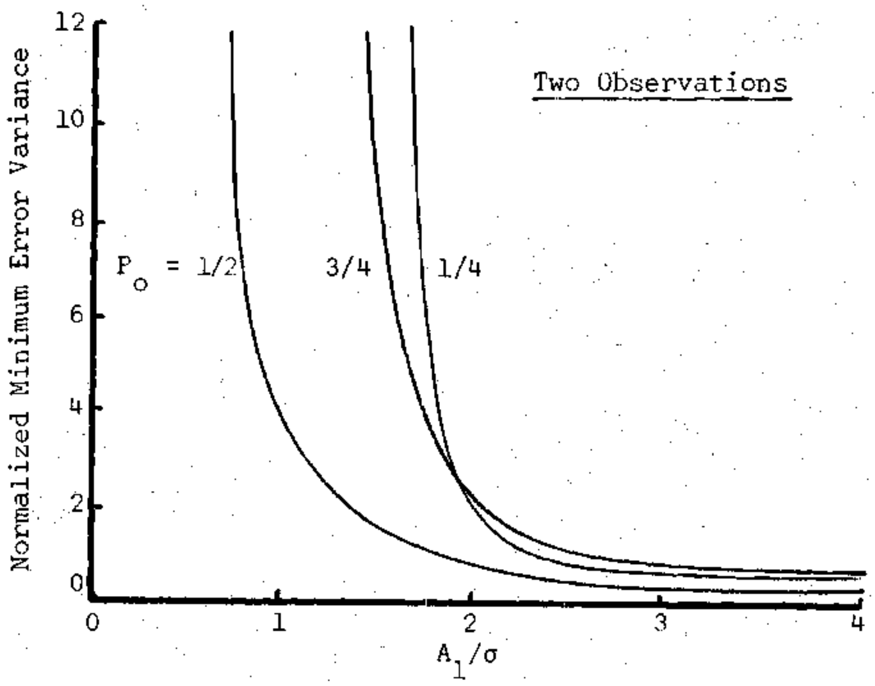


Figure 25. Minimum Error Variance for Ricean Density and the Criterion: Minimize $\{P_o\, P[1|0] + P_1\, P[0|1]\}$.

a function, $M(x)$, attains its unique maximum. Since $R(x)$ is assumed to have a unique minimum at x_o, their theorem can be applied by defining

$$M(x) = -R(x) . \tag{3-173}$$

Referring to the statement of the theorem in Appendix III, Equation (A3-1) requires that $M'(x)$ be less than zero for x greater than x_o and greater than zero for x less than x_o, for finite values of x. In terms of $R(x)$, this requires that

$$R'(x) > 0 \quad \text{for } x > x_o \tag{3-174}$$

and

$$R'(x) < 0 \quad \text{for } x < x_o , \tag{3-175}$$

for finite values of x. This is the same as the corresponding requirements of Robbins and Monro's first theorem.

Equation (A3-2) requires that the difference between $M(x)$ at any two finite values of x be bounded. Thus, $R(x)$ may have only finite discontinuities for finite values of x, but infinite discontinuities at infinity. Equation (A3-3) provides the basic requirement on defining a random variable $Y(x)$, but again offers no method for finding it. Equation (A3-4) requires that the variance of the random variable be finite.

Equations (A3-6) through (A3-8) are requirements on the rate at which the decreasing sequences, used as weighting factors in the recursive equation, approach zero.

Receiver Structure. Equations (3-174) and (3-175) are satisfied under the conditions discussed for the Robbins-Monro method; that is, if $p_o(V)$ and $p_1(V)$ are non-zero for all finite values of V. For other cases, R(x) is constant and R'(x) is zero over a range of x and some constraint is required at the receiver.

For continuous density functions, R(x) will be continuous for all values of x; if the density functions contain impulses, the discontinuities will be finite. Either type of density function will result in a R(x) which satisfies Equation (A3-2).

Define the random variable

$$Y(x) = \begin{cases} -1 \text{ if incorrect decision} \\ 0 \text{ if correct decision} \end{cases} \tag{3-176}$$

so that

$$M(x) = E[Y(x)]$$

$$M(x) = -P_o \int_x^{\infty} p_o(V)\, dV - P_1 \int_{-\infty}^{x} p_1(V)\, dV \tag{3-177}$$

and

$$M(x) = -R(x)\ . \tag{3-178}$$

$Y(x)$ takes on only finite values so that its variance is finite and Equation (A3-4) is satisfied. Then, with $\{a_n\}$ and $\{c_n\}$ defined to satisfy Equations (A3-6) through (A3-8), the recursive equation

$$x_{n+1} = x_n - \frac{a_n}{c_n}[Y(x - c_n) - Y(x + c_n)] \qquad (3\text{-}179)$$

converges to the optimum threshold, defined by Equation (3-132), with probability one as n becomes infinite.

Taking into account that the random variable for this method is the negative of the one generated for the Robbins-Monro method, Equation (3-179) is almost the same as Equation (3-143). In the earlier method, c was a positive constant and this fact resulted in the receiver converging to the wrong threshold setting. In order to obtain convergence to the true optimum threshold, c had to be allowed to approach zero which resulted in the error variance approaching infinity. The Kiefer-Wolfowitz theorem establishes the rate at which c_n may approach zero, assuring convergence to the optimum threshold, while maintaining a finite error variance.

By changing all the c inputs to c_n, the receiver of Figure 23 performs the operations required by Equation (3-179). Again, the random variables may be generated from the same observation or alternately, from independent observations.

Sacks' (21) also determined the asymptotic distribution of the error when the Kiefer-Wolfowitz process is used. This theorem is stated in Appendix IV. When one observation is used to generate both $Y(x - c_n)$ and $Y(x + c_n)$, Equation (A4-3) results in

$$\Delta^2 = 0 ,$$

or a zero error variance. However, the theorem provides the asymptotic error variance, that is, as n becomes infinite. The only conclusion can be that the theorem provides no convergence information for this case.

Convergence When Two Observations Used. Condition 1 requires that R(x) has a continuous second derivative in a neighborhood of x_o and ρ be defined as

$$\rho = \frac{R''(x_o)}{2} . \tag{3-180}$$

From Equation (3-139)

$$R''(x) = P_1\, p_1'(x) - P_o\, p_o'(x)$$

so if $p_1'(x)$ and $p_o'(x)$ exist near x_o and

$$\rho = \frac{1}{2}\,[P_1\, p_1'(x_o) - P_o\, p_o'(x_o)] \tag{3-181}$$

the condition is satisfied.

Condition 2 defines the parameter Δ^2 used in the error variance expression. Proceeding to find Δ^2,

$$\frac{\Delta^2}{2} = \lim_{x \to x_o} E\{[Y(x) - M(x)]^2\}$$

$$\frac{\Delta^2}{2} = \lim_{x \to x_o} E[Y^2(x) - 2M(x)Y(x) + M^2(x)]$$

but

$$Y^2(x) = |Y(x)| ,$$

$$E[|Y(x)|] = - M(x) = R(x) ,$$

and

$$E[Y(x)] = M(x) = - R(x)$$

so that

$$\frac{\Delta^2}{2} = \lim_{x \to x_o} \{- M(x) - 2M^2(x) + M^2(x)\}$$

$$\Delta^2 = 2[R(x_o) - R^2(x_o)] . \tag{3-182}$$

With $Y(x)$ and $M(x)$ finite for all x,

$$Z(x) = Y(x) - M(x)$$

can be bounded so that Condition 3 is always satisfied. Then, with

$$c_n = n^{-\frac{1}{4}}$$

and

$$a_n = A/n \ ,$$

where

$$A > 1/4[P_1 \, p_1'(x_o) - P_o \, p_o'(x_o)] \ ,$$

the iterative process defined by Equation (3-179) converges to the optimum threshold, x_o, in mean square and with probability one as n becomes infinite. In addition, the error, $(x_n - x_o)$, is asymptotically normal with mean zero and

$$\mathrm{Var}\ (x_n - x_o) = \frac{1}{n^{\frac{1}{2}}} \ \frac{2\ A^2\ [R(x_o) - R^2(x_o)]}{4A[P_1 \, p_1'(x_o) - P_o \, p_o'(x_o)] - 1} \ . \qquad (3\text{-}183)$$

With

$$A = 1/2[P_1 \, p_1'(x_o) - P_o \, p_o'(x_o)] \ ,$$

this expression becomes

$$\text{Min Var } (x_n - x_o) = \frac{1}{n^{\frac{1}{2}}} \; \frac{R(x_o) - R^2(x_o)}{2[P_1\, p_1'(x_o) - P_o\, p_o'(x_o)]^2} \; . \qquad (3\text{-}184)$$

This is the same as the minimum error variance, Equation (3-156), obtained using the Robbins-Monro method with c replaced by $n^{-\frac{1}{4}}$. Convergence with $n^{-\frac{1}{2}}$ is considerably slower than with n^{-1}. However, the choice is between a system which slowly converges to the true optimum threshold and a system which rapidly converges to something slightly different than the optimum threshold.

When the input is uniformly distributed, a finite value of A which satisfies the necessary conditions does not exist. Thus, the Kiefer-Wolfowitz receiver, like the Robbins-Monro receiver, may converge to the optimum threshold when the input is uniformly distributed, but the convergence properties cannot be evaluated using Sacks' theorems.

Criterion: Minimize $\{P[1|0] + P[0|1]\}$

The criterion that the sum of the conditional error probabilities be minimized has little, if any, practical application. However, it is included for completeness. This sum is given by

$$P(x) = \int_x^{\infty} p_o(V)\, dV + \int_{-\infty}^{x} p_1(V)\, dV \qquad (3\text{-}185)$$

and the optimum threshold is defined by

$$\min_{x} P(x) = P(x_o) . \qquad (3\text{-}186)$$

If $P'(x)$ is defined, the optimum threshold is also the solution to

$$p_o(x_o) = p_1(x_o) . \qquad (3\text{-}187)$$

Robbins-Monro Method

The Robbins-Monro theorem will again be applied by formulating the problem as an approximation to the derivative of the function to be minimized.

Receiver Structure When P_o Known. Define the discrete random variable

$$T(x) = \frac{1}{2c}\left[y(x + c) - y(x - c)\right]$$

where

$$y(x) = \begin{cases} \frac{1}{P_o} & \text{if "0" sent, "1" decided} \\ \frac{1}{P_1} & \text{if "1" sent, "0" decided} \\ 0 & \text{if otherwise .} \end{cases}$$

Then

$$N(x) = \frac{1}{2c}\left[\frac{1}{P_o}[P_o \int_{x+c}^{\infty} p_o(V)\, dV] + \frac{1}{P_1}[P_1 \int_{-\infty}^{x+c} p_1(V)\, dv]\right.$$

$$\left. - \frac{1}{P_o}\ [P_o \int_{x-c}^{\infty} p_o(V)\, dV] - \frac{1}{P_1}[P_1 \int_{-\infty}^{x-c} p_1(V)\, dV]\right]$$

$$N(x) = \frac{1}{2c}[P(x + c) - P(x - c)]$$

or, for small c,

$$N(x) \tilde{=} P'(x) = p_1(x) - P_o(x) \ . \qquad (3\text{-}188)$$

Since N(x) is only an approximation to P'(x), the system will converge to x_o' instead of x_o.

Differentiating N(x) gives

$$N'(x) = \frac{1}{2c}[P'(x + c) - P'(x - c)] \qquad (3\text{-}189)$$

which for small c becomes

$$N'(x) \tilde{=} P''(x) = p_1'(x) - p_o'(x) \ . \qquad (3\text{-}190)$$

For certain input probability density functions, neither the requirement that N(x) be non-decreasing nor the requirement that N(x) be non-zero for all finite values of x, other than x_o, may be satisfied. In such cases appropriate constraints, as discussed earlier, must be included in the receiver to assure convergence.

The requirement of Equation (2-12) is satisfied because T(x) takes on only finite values. Thus with $\{a_n\}$ properly defined, the process

$$x_{n+1} = x_n - \frac{a_n}{2c} [y(x + c) - y(x - c)] \qquad (3\text{-}191)$$

converges to x_o' as n becomes infinite. The receiver has the same form as the receiver of Figure 23; only P_o and P_1 inputs need be added in order to properly generate y(x) for this criterion. As before, y(x + c) and y(x - c) may be generated from the same observation or from alternate input observations.

<u>Receiver Structure When P_o Unknown.</u> Define the discrete random variable

$$T(x) = \frac{1}{2c} [y(x + c) - y(x - c)]$$

where

$$y(x) = \begin{cases} 1 \text{ if "1" sent, "0" decided and "0" sent during previous interval.} \\ 1 \text{ if "0" sent, "1" decided and "1" sent during previous interval.} \\ 0 \text{ if otherwise.} \end{cases}$$

Then

$$N(x) = \frac{1}{2c}\left[P_o(P_1 \int_{-\infty}^{x+c} p_1(V)\,dV) + P_1(P_o \int_{x+c}^{\infty} p_o(V)\,dV)\right.$$

$$\left. - [P_o(P_1 \int_{-\infty}^{x-c} p_1(V)\,dV) + P_1(P_o \int_{x-c}^{\infty} p_o(V)\,dV)]\right]$$

or

$$N(x) = \frac{P_o\, P_1}{2c}\,[P(x + c) - P(x - c)]$$

so that for small c,

$$N(x) \cong P_o\, P_1\, P'(x) = P_o\, P_1[p_1(x) - p_o(x)]\ . \qquad (3\text{-}192)$$

Differentiating N(x)

$$N'(x) = \frac{P_o\, P_1}{2c}\,[P'(x + c) - P'(x - c)] \qquad (3\text{-}193)$$

which for small c becomes

$$N'(x) \cong P_o\, P_1\, P''(x) = P_o\, P_1[p_1'(x) - p_o'(x)]\ . \qquad (3\text{-}194)$$

Then with $\{a_n\}$ properly defined, and taking into account any required receiver constraints, the recursive process

$$x_{n+1} = x_n - \frac{a_n}{2c}[y(x + c) - y(x - c)] \tag{3-195}$$

converges to x_o' as n becomes infinite. This receiver is the same as the one of the Figure 23 except that $y(x + c)$ and $y(x - c)$ depend on the previous signal in addition to the present signal and decision.

<u>Convergence When P_o Known and One Observation Used.</u> Condition 1 is satisfied by the problem formulation and the inclusion of any receiver constraints required. Conditions 2, 4(a) and 5 are satisfied because $T(x)$ and $N(x)$ are finite for all x. Condition 3 is satisfied if c is small and

$$\alpha_1 = p_1'(x_o') - p_o'(x_o') . \tag{3-196}$$

Then

$$T^2(x) = \frac{1}{4c^2}[y^2(x + c) - 2y(x + c)y(x - c) + y^2(x - c)]$$

where, when one observation is used to generate $y(x + c)$ and $y(x - c)$,

$$y(x + c)y(x - c) = \begin{cases} \dfrac{1}{P_o^2} & \text{if "0" sent and received signal above } x + c \\ \dfrac{1}{P_1^2} & \text{if "1" sent and received signal below } x - c \\ 0 & \text{if otherwise} \end{cases}$$

so that

$$\gamma^2 = \frac{1}{4c^2}\left[\frac{1}{P_o^2}\left(P_o \int_{x_o'+c}^{\infty} p_o(V)\, dV\right) + \frac{1}{P_1^2}\left(P_1 \int_{-\infty}^{x_o'+c} p_1(V)\, dV\right)\right.$$

$$- 2\left[\frac{1}{P_o^2}\left(P_o \int_{x_o'+c}^{\infty} p_o(V)\, dV\right) + \frac{1}{P_1^2}\left(P_1 \int_{-\infty}^{x_o'-c} p_1(V)\, dV\right)\right]$$

$$\left. + \frac{1}{P_o^2}\left(P_o \int_{x_o'-c}^{\infty} p_o(V)\, dV\right) + \frac{1}{P_1^2}\left(P_1 \int_{-\infty}^{x_o'-c} p_1(V)\, dV\right)\right] .$$

Combining terms, this becomes

$$\gamma^2 = \frac{1}{4c^2}\left[- \frac{1}{P_o}\int_{x_o'+c}^{\infty} p_o(V)\, dV + \frac{1}{P_o}\int_{x_o'-c}^{\infty} p_o(V)\, dV\right.$$

$$\left. + \frac{1}{P_1}\int_{-\infty}^{x_o'+c} p_1(V)\, dV - \frac{1}{P_1}\int_{-\infty}^{x_o'-c} p_1(V)\, dV\right]$$

which, for small c, becomes

$$\gamma^2 \cong \frac{1}{2c}\left[\frac{1}{P_o}\, p_o(x_o') + \frac{1}{P_1}\, p_1(x_o')\right] .$$

But from the criterion, and for small c,

$$p_o(x_o') \cong p_1(x_o')$$

so that

$$\gamma^2 \cong \frac{1}{2c} \frac{p_o(x_o')}{P_o P_1} . \tag{3-197}$$

With

$$a_n = A/n ,$$

where

$$A > 1/2[p_1'(x_o') - p_o'(x_o')] ,$$

$$\mathrm{Var}\,(x_n - x_o') \cong \frac{1}{2ncP_oP_1} \frac{A^2\, p_o(x_o')}{2A[p_1'(x_o') - p_o'(x_o')] - 1} \tag{3-198}$$

or with

$$A = 1/[p_1'(x_o') - p_o'(x_o')] ,$$

$$\mathrm{Min\ Var}\,(x_n - x_o') \cong \frac{1}{nc} \frac{p_o(x_o')}{2P_oP_1[p_1'(x_o') - p_o'(x_o')]^2} . \tag{3-199}$$

This expression results in a curve with the same shape as the P_o equal one-half curve obtained from Equation (3-153); only the amplitude need be scaled by $1/4P_oP_1$ for the a priori probabilities of interest.

Convergence When P_o Known and Two Observations Used. When separate, independent observations are used to generate $y(x + c)$ and $y(x - c)$,

$$E[y(x + c)y(x - c)] = E[y(x + c)]E[y(x - c)] ,$$

$$E[y(x + c)] = P(x + c) ,$$

$$E[y(x - c)] = P(x - c) ,$$

and from the formulation of the problem,

$$P(x_o' + c) = P(x_o' - c) .$$

Then

$$E[T^2(x)] = \frac{1}{4c^2}\left[\frac{1}{P_o^2}\left[P_o\int_{x+c}^{\infty} p_o(V)\ dV\right] + \frac{1}{P_1^2}\left[P_1\int_{-\infty}^{x+c} p_1(V)\ dV\right]\right.$$

$$- 2P(x + c)P(x - c)$$

$$\left. + \frac{1}{P_o^2}\left[P_o\int_{x-c}^{\infty} p_o(V)\ dV\right] + \frac{1}{P_1^2}\left[P_1\int_{-\infty}^{x-c} p_1(V)\ dV\right]\right]$$

so that

$$\gamma^2 = \frac{1}{4c^2}\left[\frac{1}{P_o}\left[\int_{x_o'+c}^{\infty} p_o(V)\,dV + \int_{x_o'-c}^{\infty} p_o(V)\,dV\right]\right.$$

$$+ \frac{1}{P_1}\left[\int_{-\infty}^{x_o'+c} p_1(V)\,dV + \int_{-\infty}^{x_o'-c} p_1(V)\,dV\right]$$

$$\left. - 2\,P^2(x_o' + c)\right]$$

which, after some manipulation and assuming c small, becomes

$$\gamma^2 \cong \frac{1}{2c^2 P_o P_1}\,[P(x_o') - R(x_o') - P_o P_1\, P^2(x_o')] \ , \qquad (3\text{-}200)$$

where R(x) is the average probability of error defined by Equation (3-131). Then with

$$a_n = A/n \ ,$$

where

$$A > 1/2[p_1'(x_o') - p_o'(x_o')] \ ,$$

$$\mathrm{Var}\,(x_n - x_o') \cong \frac{1}{2nc^2 P_o P_1} \, \frac{A^2[P(x_o') - R(x_o') - P_o P_1 P^2(x_o')]}{2\,A[p_1'(x_o') - p_o'(x_o')] - 1} \quad (3\text{-}201)$$

and when

$$A = 1/[p_1'(x_o') - p_o'(x_o')] \ ,$$

$$\mathrm{Min\ Var}\,(x_n - x_o') \cong \frac{1}{nc^2} \, \frac{P(x_o') - R(x_o') - P_o P_1 P^2(x_o')}{2P_o P_1 [p_1'(x_o') - p_o'(x_o')]^2} \ . \qquad (3\text{-}202)$$

This expression results in a curve similar to the P_o equal one-half curve obtained from Equation (3-156), especially when the "1" signal has a large amplitude and the second order terms of the numerators can be neglected.

Convergence When P_o Unknown and One Observation Used. When the a priori probabilities are unknown and one observation is used to generate both random variables,

$$y(x + c)y(x - c) = \begin{cases} 1 \text{ if "1" sent, received signal below } x - c \\ \quad \text{and "0" sent during previous interval.} \\ 1 \text{ if "0" sent, received signal above } x + c \\ \quad \text{and "1" sent during previous interval.} \\ 0 \text{ if otherwise} \end{cases}$$

so that

$$E[y(x + c)y(x - c)] = P_o[P_1 \int_{-\infty}^{x-c} p_1(V)\ dV] + P_1[P_o \int_{x+c}^{\infty} p_o(V)\ dV] .$$

In addition

$$y^2(x + c) = y(x + c) ,$$

$$y^2(x - c) = y(x - c) ,$$

and

$$\lim_{x \to x_o'} E[y(x + c)] = \lim_{x \to x_o'} E[y(x - c)]$$

from the formulation, so that

$$\gamma^2 = \frac{1}{4c^2}\left[2\left[P_o\left(P_1 \int_{-\infty}^{x_o'+c} p_1(V)\ dV\right) + P_1\left(P_o \int_{x_o'+c}^{\infty} p_o(V)\ dV\right)\right]\right.$$

$$\left. -2\left[P_o\left(P_1 \int_{-\infty}^{x_o'-c} p_1(V)\ dV\right) + P_1\left(P_o \int_{x_o'+c}^{\infty} p_o(V)\ dV\right)\right]\right]$$

or

$$\gamma^2 = \frac{1}{2c^2} P_o P_1\left[\int_{-\infty}^{x_o'+c} p_1(V)\ dV - \int_{-\infty}^{x_o'-c} p_1(V)\ dV\right]$$

or, for small c,

$$\gamma^2 \cong \frac{P_o P_1}{c} p_1(x_o') \cong \frac{P_o P_1}{c} p_o(x_o') \ . \qquad (3\text{-}203)$$

With

$$A > 1/2P_o P_1[p_1'(x_o') - p_o'(x_o')] \ ,$$

$$\text{Var}\ (x_n - x_o') \cong \frac{1}{nc} \frac{P_o\ P_1\ p_o(x_o')\ A^2}{2AP_o P_1[p_1'(x_o') - p_o'(x_o')] - 1} \qquad (3\text{-}204)$$

and with

$$A = 1/P_o P_1[p_1'(x_o') - p_o'(x_o')] \ ,$$

$$\text{Min Var}\ (x_n - x_o') \cong \frac{1}{nc} \frac{p_o(x_o')}{P_o P_1[p_1'(x_o') - p_o'(x_o')]^2} \ . \qquad (3\text{-}205)$$

This is just twice the value of the corresponding minimum error variance, Equation (3-199), when the a priori probabilities are known.

Convergence when P_o Unknown and Two Observations Used. When separate, independent observations are used to generate the two random variables, and c is small,

$$\lim_{x \to x_o'} E[y(x + c)y(x - c)] = [P_o P_1\ P(x_o')]^2$$

so that

$$\gamma^2 = \frac{1}{2c^2}\{P_o P_1\ P(x_o') - [P_o P_1\ P(x_o')]^2\}\ . \qquad (3\text{-}206)$$

Therefore, with

$$a_n = A/n\ ,$$

where

$$A > 1/2P_o P_1[p_1'(x_o') - p_o'(x_o')]\ ,$$

$$\mathrm{Var}\ (x_n - x_o') \cong \frac{1}{2nc^2}\,\frac{P_o P_1[P(x_o') - P_o P_1\ P^2(x_o')]A^2}{2P_o P_1[p_1'(x_o') - p_o'(x_o')]A - 1} \qquad (3\text{-}207)$$

or, with

$$A = 1/P_o P_1[p_1'(x_o') - p_o'(x_o')]\ ,$$

$$\mathrm{Min\ Var}\ (x_n - x_o') \cong \frac{1}{nc^2}\ \frac{P(x_o') - P_o P_1\ P^2(x_o')}{2P_o P_1[p_1'(x_o') - p_o'(x_o')]^2}\ . \qquad (3\text{-}208)$$

This expression is only slightly greater than the corresponding expression when the a priori probabilities are known, Equation (3-202), due to the extra negative term in the numerator of the earlier equation.

Kiefer-Wolfowitz Method

Receiver Structure when P_o Known. Define the random variable

$$Y(x) = \begin{cases} -\frac{1}{P_o} \text{ if "0" sent, "1" decided.} \\ -\frac{1}{P_1} \text{ if "1" sent, "0" decided.} \\ 0 \text{ if otherwise.} \end{cases}$$

so that

$$M(x) = -\frac{1}{P_o}[P_o \int_x^\infty p_o(V)\, dV] - \frac{1}{P_1}[P_1 \int_{-\infty}^x p_1(V)\, dV] \quad (3\text{-}209)$$

or

$$M(x) = -P(x) . \quad (3\text{-}210)$$

Then with $\{a_n\}$ and $\{c_n\}$ properly defined, and any required constraints included in the receiver, all of the conditions are satisfied and

$$x_{n+1} = x_n - \frac{a_n}{c_n}[Y(x - c_n) - Y(x + c_n)] \quad (3\text{-}211)$$

converges to x_o with probability one as n becomes infinite. This receiver is the same as the one derived using the Robbins-Monro method except that a decreasing c is used, resulting in convergence to the true optimum threshold.

Receiver Structure when P_o Unknown. Define the random variable

$$Y(x) = \begin{cases} -1 \text{ if "1" sent, "0" decided and "0"} \\ \quad\quad \text{sent during previous interval.} \\ -1 \text{ if "0" sent, "1" decided and "1"} \\ \quad\quad \text{sent during previous interval.} \\ 0 \text{ if otherwise.} \end{cases}$$

so that

$$M(x) = - P_o[P_1 \int_{-\infty}^{x} p_1(V)\, dV] - P_1[P_o \int_{x}^{\infty} p_o(V)\, dV] \qquad (3\text{-}212)$$

or

$$M(x) = - P_o P_1 \, P(x) \, . \qquad (3\text{-}213)$$

Again, all of the conditions are satisfied and

$$x_{n+1} = x_n - \frac{a_n}{c_n} [Y(x - c_n) - Y(x + c_n)] \qquad (3\text{-}214)$$

converges to the optimum threshold.

Convergence when P_o Known and Two Observations Used. Differentiating Equation (3-210) twice gives

$$M''(x) = - P''(x) \qquad (3\text{-}215)$$

or, from the definition of P(x),

$$M''(x) = p_o'(x) - p_1'(x) \; . \tag{3-216}$$

Condition 1 is therefore satisfied if $p_o'(x)$ and $p_1'(x)$ are continuous in a neighborhood of x_o and ρ is defined as

$$\rho = \frac{1}{2}[p_1'(x_o) - p_o'(x_o)] \; . \tag{3-217}$$

Since $Y(x)$ is finite for all values of x, Condition 3 is satisfied. Then, by Condition 2,

$$\frac{\Delta^2}{2} = \lim_{x \to x_o} \{E[Y^2(x)] - 2M^2(x) + M^2(x)\}$$

$$\frac{\Delta^2}{2} = \lim_{x \to x_o} \left[\frac{1}{P_o^2}\left[P_o \int_x^\infty p_o(V)\, dV\right] + \frac{1}{P_1^2}\left[P_1 \int_{-\infty}^x p_1(V)\, dV\right] - M^2(x)\right]$$

$$\frac{\Delta^2}{2} = \frac{1}{P_o P_1}[P_1 \int_{x_o}^\infty p_o(V)\, dV + P_o \int_{-\infty}^{x_o} p_1(V)\, dV] - P^2(x_o)$$

which can be written as

$$\Delta^2 = \frac{2}{P_o P_1}[P(x_o) - R(x_o) - P_o P_1\, P^2(x_o)] \; . \tag{3-218}$$

Then with

$$c_n = n^{-\frac{1}{4}} ,$$

$$a_n = A/n ,$$

where

$$A > 1/4[p_1'(x_o) - p_o'(x_o)] ,$$

the iterative process defined by Equation (3-211) converges to x_o in mean square and with probability one as n becomes infinite, and

$$\mathrm{Var}\ (x_n - x_o) = \frac{1}{n^{\frac{1}{2}} P_o P_1} \; \frac{2[P(x_o) - R(x_o) - P_o P_1 P^2(x_o)]A^2}{4A\ [p_1'(x_o) - p_o'(x_o)] - 1} . \quad (3\text{-}219)$$

If

$$A = 1/2[p_1'(x_o) - p_o'(x_o)] ,$$

$$\mathrm{Min\ Var}\ (x_n - x_o) = \frac{1}{n^{\frac{1}{2}}} \; \frac{P(x_o) - R(x_o) - P_o P_1 P^2(x_o)}{2P_o P_1 [p_1'(x_o) - p_o'(x_o)]^2} . \quad (3\text{-}220)$$

This is similar to the corresponding expression when the Robbins-Monro method is used, Equation (3-202). The difference is that the Kiefer-Wolfowitz receiver converges like $1/n^{\frac{1}{2}}$ whereas the Robbins-Monro receiver converges like $1/n$.

Convergence When P_o Unknown and Two Observations Used.

Differentiating Equation (3-213) twice gives

$$M''(x) = - P_oP_1 \, P''(x) \tag{3-221}$$

$$M''(x) = P_oP_1[p_o'(x) - p_1'(x)] \tag{3-222}$$

so that Condition 1 is satisfied if $p_o'(x)$ and $p_1'(x)$ are continuous in a neighborhood of x_o and ρ is defined as

$$\rho = \frac{1}{2} P_oP_1[p_1'(x_o) - p_o'(x_o)] \; . \tag{3-223}$$

For this case

$$Y^2(x) = - Y(x)$$

so that

$$\frac{\Delta^2}{2} = \lim_{x \to x_o} [-M(x) - 2M^2(x) + M^2(x)]$$

$$\Delta^2 = 2\{P_oP_1 \, P(x_o) - [P_oP_1 \, P(x_o)]^2\} \; . \tag{3-224}$$

Then, with

$$c_n = n^{-\frac{1}{4}} ,$$

$$a_n = A/n ,$$

where

$$A > 1/4P_o P_1[p_1'(x_o) - p_o'(x_o)] ,$$

the iterative process defined by Equation (3-214) converges to x_o in mean square and with probability one as n becomes infinite, and

$$\mathrm{Var}\ (x_n - x_o) = \frac{1}{\frac{1}{n^2}} \frac{2P_o P_1[P(x_o) - P_o P_1\ P^2(x_o)]A^2}{4P_o P_1 A[p_1'(x_o) - p_o'(x_o)] - 1} . \qquad (3\text{-}225)$$

If

$$A = 1/2P_o P_1[p_1'(x_o) - p_o'(x_o)] ,$$

$$\mathrm{Min\ Var}\ (x_n - x_o) = \frac{1}{\frac{1}{n^2}} \frac{P(x_o) - P_o P_1\ P^2(x_o)}{2P_o P_1[p_1'(x_o) - p_o'(x_o)]^2} , \qquad (3\text{-}226)$$

which is similar to the corresponding expression when the Robbins-Monro method is used, Equation (3-208), except for the slower rate of convergence.

CHAPTER IV

MULTIPLE THRESHOLD ADAPTIVE RECEIVERS

Previous chapters considered receivers which were required to decide between only two possible transmitted symbols. Thus, only one threshold was needed. However, in more advanced communication systems, the transmitter may select any one of M symbols for transmission during each signal interval, so that M - 1 receiver thresholds are required. For a given criterion of performance for such a M-ary system, the set of M - 1 optimum threshold settings is defined. If the receiver signal statistics are known, the optimum settings can be determined using the standard techniques discussed earlier. However, if all of the statistical data is not available, a multidimensional adaptive or learning receiver is required.

Blum (22) extended the Robbins-Monro method to the multidimensional process for determining the vector $\underline{x}_o$,

$$\underline{x}_o = \begin{bmatrix} x_{1o} \\ x_{2o} \\ \cdot \\ \cdot \\ \cdot \\ x_{qo} \end{bmatrix}, \tag{4-1}$$

for which the vector valued function,

$$\underline{N}(\underline{x}) = \begin{bmatrix} N_1(\underline{x}) \\ N_2(\underline{x}) \\ \cdot \\ \cdot \\ \cdot \\ N_q(\underline{x}) \end{bmatrix}, \tag{4-2}$$

where

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_q \end{bmatrix}, \tag{4-3}$$

is zero. Thus

$$\underline{N}(\underline{x}_o) = \underline{0} \tag{4-4}$$

or

$$\begin{aligned} N_1(x_{1o}, x_{2o}, \ldots, x_{qo}) &= 0 \\ N_2(x_{1o}, x_{2o}, \ldots, x_{qo}) &= 0 \\ &\vdots \\ N_q(x_{1o}, x_{2o}, \ldots, x_{qo}) &= 0 \,. \end{aligned} \tag{4-5}$$

Blum (23) also extended the Kiefer-Wolfowitz technique to the multidimensional process for determining the vector $\underset{\sim}{x}_o$ for which the scalar valued function

$$f(\underline{x}) = f(x_1, x_2, \ldots , x_q) \tag{4-6}$$

attains its maximum.

Sacks (24), (25) later determined the asymptotic distributions of the error for both processes. His theorems, which include both the conditions under which the processes will converge and the conditions under which the asymptotic distributions can be determined, are given in Appendices V and VI.

To demonstrate and evaluate the multidimensional techniques, receivers for two criteria of engineering interest are derived and their convergence properties determined. It is assumed that the transmitter selects one of three possible symbols during each signal interval, so that two decision thresholds are required at the receiver.

<u>Criterion: $P_o\ P[\varepsilon|0] = P_1\ P[\varepsilon|1] = P_2\ P[\varepsilon|2]$</u>

When the criterion requires that the probabilities of sending a "0" and making a decision error, sending a "1" and making a decision error and sending a "2" and making a decision error, are equal, the decision thresholds are defined by

$$P_o \int_{x_{1o}}^{\infty} p_o(V)\, dV = P_1 \int_{-\infty}^{x_{1o}} p_1(V)\, dV + P_1 \int_{x_{2o}}^{\infty} p_1(V)\, dV = P_2 \int_{-\infty}^{x_{2o}} p_2(V)\, dV \ . \tag{4-7}$$

A typical solution is shown in Figure 26.

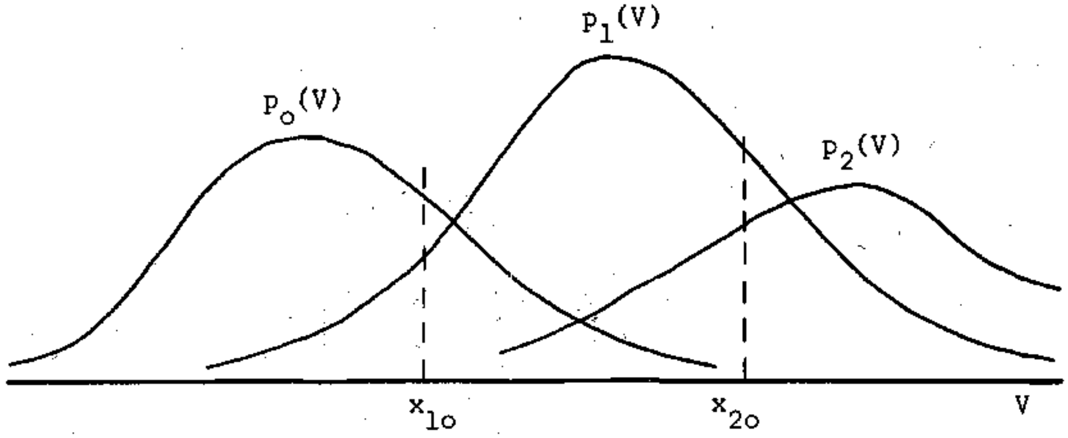


Figure 26. Typical Input Density Functions and Optimum Thresholds

Receiver Structure

In order to employ the multidimensional Robbins-Monro method, two random variables, $T_1(\underline{x})$ and $T_2(\underline{x})$, must be generated. Define

$$T_1(\underline{x}) = \begin{cases} 1 \text{ if "1" sent, "0" or "2" decided.} \\ -1 \text{ if "0" sent, "1" or "2" decided.} \\ 0 \text{ if otherwise.} \end{cases} \tag{4-8}$$

and

$$T_2(\underline{x}) = \begin{cases} 1 \text{ if "2" sent, "0" or "1" decided.} \\ -1 \text{ if "1" sent, "0" or "2" decided.} \\ 0 \text{ if otherwise.} \end{cases} \tag{4-9}$$

so that

$$N_1(\underline{x}) = E[T_1(\underline{x})]$$

$$N_1(\underline{x}) = P_1 \int_{-\infty}^{x_1} p_1(V)\, dV + P_1 \int_{x_2}^{\infty} p_1(V)\, dV - P_o \int_{x_1}^{\infty} p_o(V)\, dV \quad (4\text{-}10)$$

and

$$N_2(\underline{x}) = E[T_2(\underline{x})]$$

$$N_2(\underline{x}) = P_2 \int_{-\infty}^{x_2} p_2(V)\, dV - P_1 \int_{-\infty}^{x_1} p_1(V)\, dV - P_1 \int_{x_2}^{\infty} p_1(V)\, dV. \quad (4\text{-}11)$$

Evaluating at $\underline{x}_o$ gives

$$N_1(\underline{x}_o) = P_1\, P[\varepsilon|1] - P_o\, P[\varepsilon|0] = 0$$

and

$$N_2(\underline{x}_o) = P_2\, P[\varepsilon|2] - P_1\, P[\varepsilon|1] = 0$$

and Equations (A5-1) and (A5-2) are satisfied. Then, with $\{a_n\}$ defined to satisfy Equations (A5-3) and (A5-4), the thresholds during training are adjusted according to

$$x_{1,n+1} = x_{1,n} - a_n\, T_1(\underline{x}) \quad (4\text{-}12)$$

and

$$x_{2,n+1} = x_{2,n} - a_n \, T_2(\underline{x}) \; . \tag{4-13}$$

These thresholds will converge to the optimum thresholds, x_{1o} and x_{2o}, with probability one as n approaches infinity if Conditions 1, 2, and 3 are satisfied.

Condition 1 is analogous to the one dimensional requirement that N(x) be positive for finite values of x greater than x_o and negative for finite values of x less than x_o. However, both the interpretation of the condition and the proof that the condition is satisfied are more difficult in the multidimensional case. Let

$$g \triangleq (\underline{x} - \underline{x}_o, \underline{N}(\underline{x})) \; ,$$

then

$$g = (x_1 - x_{1o})N_1(\underline{x}) + (x_2 - x_{2o})N_2(\underline{x})$$

$$g = (x_1 - x_{1o})[P_1 \int_{-\infty}^{x_1} p_1(V)\, dV + P_1 \int_{x_2}^{\infty} p_1(V)\, dV - P_o \int_{x_1}^{\infty} p_o(V)\, dV]$$

$$+ (x_2 - x_{2o})[P_2 \int_{-\infty}^{x_2} p_2(V)\, dV - P_1 \int_{-\infty}^{x_1} p_1(V)\, dV - P_1 \int_{x_2}^{\infty} p_1(V)\, dV] \; .$$

This may be written as

$$g = [(x_1 - x_{1o}) - (x_2 - x_{2o})]\, P_1 \left[\int_{x_{1o}}^{x_1} p_1(V)\, dV - \int_{x_{2o}}^{x_2} p_1(V)\, dV \right] \quad (4\text{-}14)$$

$$+ (x_1 - x_{1o})\, P_o \int_{x_{1o}}^{x_1} p_o(V)\, dV + (x_2 - x_{2o})\, P_2 \int_{x_{2o}}^{x_2} p_2(V)\, dV \ .$$

Condition 1 will be satisfied if g is positive for all $\underline{x}$ such that

$$0 < \|\underline{x} - \underline{x}_o\| < \infty \ .$$

The last two terms are always positive for

$$x_1 \neq x_{1o} \text{ and } x_2 \neq x_{2o} \ .$$

The first term is zero for

$$x_1 - x_{1o} = x_2 - x_{2o}$$

and positive for

$$x_1 - x_{1o} < 0 < x_2 - x_{2o}$$

and

$$x_2 - x_{2o} < 0 < x_1 - x_{1o}$$

regardless of the density functions and a priori probabilities. Thus, for these combinations of x_1 and x_2, g is positive. For other choices of x_1 and x_2, the first term may be negative and the shape of the density functions and the a priori probabilities will determine if the last two terms are large enough to keep g positive. The functional form of g makes an explicit determination of conditions on the density functions and a priori probabilities difficult. However, examination of Equation (4-14), based on Figure 26, indicates that the condition is satisfied if the density functions are unimodal and the a priori probabilities are such that x_{1o} and x_{2o} lie between the peaks. For this case any decrease in the first term is offset by a larger increase in the last two terms. Thus, the two important cases of gaussian and Ricean noise, with a reasonable amplitude separation between signals, will satisfy the condition. Other types of input statistics may or may not satisfy the condition.

Condition 2 requires that there exists a K_1 such that

$$N_1^2(\underline{x}) + N_2^2(\underline{x}) \le K_1(x_1 - x_{1o})^2 + K_1(x_2 - x_{2o})^2 \, , \qquad (4\text{-}15)$$

that is, $N_1(\underline{x})$ and $N_2(\underline{x})$ are finite for x_1 and x_2 finite. Since $N_1(\underline{x})$ and $N_2(\underline{x})$ are finite for all $\underline{x}$, this condition is always satisfied.

Condition 3 requires that

$$\sup_{\underline{x}} E\{[T_1(\underline{x}) - N_1(\underline{x})]^2 + [T_2(\underline{x}) - N_2(\underline{x})]^2\} < \infty \, . \qquad (4\text{-}16)$$

Since $T_1(\underline{x})$ and $T_2(\underline{x})$ take on only finite values, their variances are finite for all $\underline{x}$ so that the sum of their variance is finite for all $\underline{x}$ and the condition is always satisfied.

Therefore, if the input statistics are such that g is positive, all of the conditions for convergence are satisfied and the receiver based on Equations (4-12) and (4-13) converges to the optimum thresholds. A block diagram of the receiver is given in Figure 27.

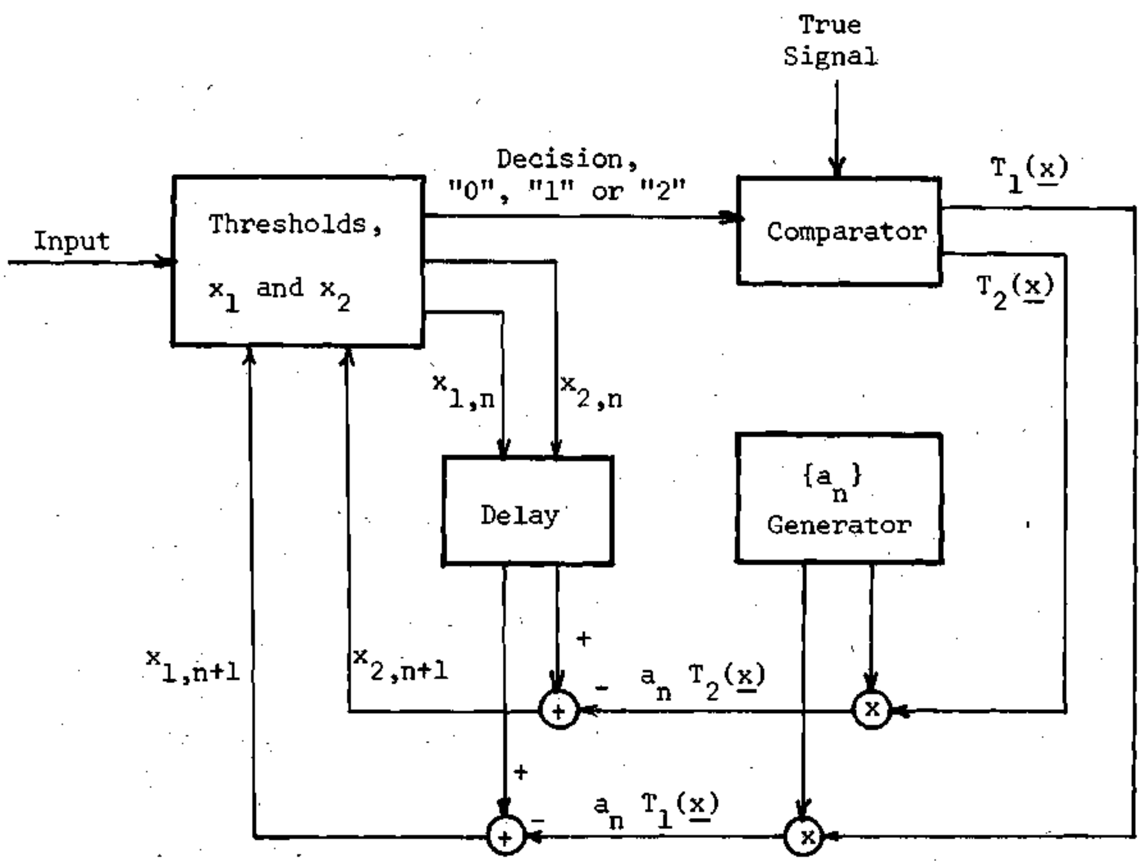


Figure 27. Receiver Based on the Multidimensional Robbins-Monro Technique

Convergence

Since $\underline{T}(\underline{x})$ takes on only finite values, Condition 6 is always satisfied. Condition 4 defines

$$\underline{\pi} = \lim_{\underline{x}\to\underline{x}_o} E(\underline{Z}(\underline{x})\ \underline{Z}^T(\underline{x}))$$

where

$$\underline{Z}(\underline{x}) = \underline{T}(\underline{x}) - \underline{N}(\underline{x}) .$$

But noting that

$$\underline{N}(\underline{x}_o) = \underline{0} ,$$

$\underline{\pi}$ may be determined from

$$\underline{\pi} = \lim_{\underline{x}\to\underline{x}_o} E(\underline{T}(\underline{x})\ \underline{T}^T(\underline{x})) . \tag{4-17}$$

Then

$$\underline{T}(\underline{x})\ \underline{T}^T(\underline{x}) = \begin{bmatrix} T_1^{\,2}(\underline{x}) & T_1(\underline{x})\ T_2(\underline{x}) \\ T_1(\underline{x})\ T_2(\underline{x}) & T_2^{\,2}(\underline{x}) \end{bmatrix}$$

where

$$T_1^{\,2}(\underline{x}) = \begin{cases} 1 \text{ if "0" or "1" sent and incorrect decision made.} \\ 0 \text{ if otherwise,} \end{cases}$$

$$T_2^{\,2}(\underline{x}) = \begin{cases} 1 \text{ if "1" or "2" sent and incorrect decision made.} \\ 0 \text{ if otherwise,} \end{cases}$$

and

$$T_1(\underline{x})\ T_2(\underline{x}) = \begin{cases} -1 \text{ if "1" sent and "0" or "2" decided.} \\ 0 \text{ if otherwise.} \end{cases}$$

so that

$$E[T_1^{\,2}(\underline{x})] = P_o \int_{x_1}^{\infty} p_o(V)\ dV + P_1 \int_{-\infty}^{x_1} p_1(V)\ dV + P_1 \int_{x_2}^{\infty} p_1(V)\ dV$$

and

$$\pi_{11} = \lim_{\underline{x} \to \underline{x}_o} E[T_1^{\,2}(\underline{x})] \ ,$$

$$\pi_{11} = P_o \int_{x_{1o}}^{\infty} p_o(V)\ dV + P_1 \int_{-\infty}^{x_{1o}} p_1(V)\ dV + P_1 \int_{x_{2o}}^{\infty} p_1(V)\ dV$$

or, using one of the equalities of the criterion,

$$\pi_{11} = 2\ P_o \int_{x_{1o}}^{\infty} p_o(V)\ dV \ . \tag{4-18}$$

Similarly

$$E[T_2^{\,2}(\underline{x})] = P_1 \int_{-\infty}^{x_1} p_1(V)\ dV + P_1 \int_{x_2}^{\infty} p_1(V)\ dV + P_2 \int_{-\infty}^{x_2} p_2(V)\ dV$$

and

$$\pi_{22} = 2\, P_o \int_{x_{1o}}^{\infty} p_o(V)\, dV \, , \tag{4-19}$$

and

$$E[T_1(\underline{x})\, T_2(\underline{x})] = -\, P_1 \int_{-\infty}^{x_1} p_1(V)\, dV - P_1 \int_{x_2}^{\infty} p_1(V)\, dV$$

so that

$$\pi_{12} = \pi_{21} = -\, P_o \int_{x_{1o}}^{\infty} p_o(V)\, dV \, . \tag{4-20}$$

Then with

$$\underline{\pi} = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{12} & \pi_{11} \end{bmatrix} \tag{4-21}$$

and the orthogonal matrix

$$\underline{P} = \begin{bmatrix} -1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} , \tag{4-22}$$

$$\underline{P}^{-1}\, \underline{\pi}\, \underline{P} = \begin{bmatrix} \pi_{11} - \pi_{12} & 0 \\ 0 & \pi_{11} + \pi_{12} \end{bmatrix} \tag{4-23}$$

$$\underline{P}^{-1}\,\underline{\pi}\,\underline{P} = \begin{bmatrix} 3\,P_o \int_{x_{1o}}^{\infty} p_o(V)\,dV & 0 \\ 0 & P_o \int_{x_{1o}}^{\infty} p_o(V)\,dV \end{bmatrix} . \qquad (4\text{-}24)$$

Since the diagonal elements, which are also the characteristic roots, are positive, $\underline{\pi}$ is positive definite and thus satisfies the requirement that it be non-negative definite.

Writing $\underline{N}(\underline{x})$ in the form required by Condition 5,

$$\underline{N}(\underline{x}) = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} x_1 - x_{1o} \\ x_2 - x_{2o} \end{bmatrix} + \underline{\delta}(\underline{x},\underline{x}_o) \qquad (4\text{-}25)$$

so that

$$N_1(\underline{x}) = b_{11}(x_1 - x_{1o}) + b_{12}(x_2 - x_{2o}) + \delta_1(\underline{x},\underline{x}_o) \quad . \qquad (4\text{-}26)$$

and

$$N_2(\underline{x}) = b_{21}(x_1 - x_{1o}) + b_{22}(x_2 - x_{2o}) + \delta_2(\underline{x},\underline{x}_o) \ . \qquad (4\text{-}27)$$

Differentiating Equation (4-26) with respect to x_1 results in

$$\frac{\partial N_1(\underline{x})}{\partial x_1} = b_{11} + \frac{\partial \delta_1(\underline{x},\underline{x}_o)}{\partial x_1} \tag{4-28}$$

and evaluating at $\underline{x}_o$,

$$\left.\frac{\partial N_1(\underline{x})}{\partial x_1}\right|_{\underline{x}=\underline{x}_o} = b_{11} + \left.\frac{\partial \delta_1(\underline{x},\underline{x}_o)}{\partial x_1}\right|_{\underline{x}=\underline{x}_o} . \tag{4-29}$$

But, by virtue of Equation (A5-12),

$$\left.\frac{\partial \delta_1(\underline{x},\underline{x}_o)}{\partial x_1}\right|_{\underline{x}=\underline{x}_o} = \left.\frac{\partial \delta_1(\underline{x},\underline{x}_o)}{\partial x_2}\right|_{\underline{x}=\underline{x}_o} = 0 \tag{4-30}$$

and

$$\left.\frac{\partial \delta_2(\underline{x},\underline{x}_o)}{\partial x_1}\right|_{\underline{x}=\underline{x}_o} = \left.\frac{\partial \delta_2(\underline{x},\underline{x}_o)}{\partial x_2}\right|_{\underline{x}=\underline{x}_o} = 0$$

so that

$$\left.\frac{\partial N_1(\underline{x})}{\partial x_1}\right|_{\underline{x}=\underline{x}_o} = b_{11} . \tag{4-31}$$

Similarly

$$\left.\frac{\partial\ N_1(\underline{x})}{\partial\ x_2}\right|_{\underline{x}=\underline{x}_o} = b_{12}\ , \tag{4-32}$$

$$\left.\frac{\partial\ N_2(\underline{x})}{\partial\ x_1}\right|_{\underline{x}=\underline{x}_o} = b_{21}\ , \tag{4-33}$$

$$\left.\frac{\partial\ N_2(\underline{x})}{\partial\ x_2}\right|_{\underline{x}=\underline{x}_o} = b_{22}\ . \tag{4-34}$$

From the definitions of $N_1(\underline{x})$ and $N_2(\underline{x})$, Equations (4-10) and (4-11),

$$b_{11} = P_1\ p_1(x_{1o}) + P_o\ p_o(x_{1o})\ , \tag{4-35}$$

$$b_{12} = -\ P_1\ p_1(x_{2o})\ , \tag{4-36}$$

$$b_{21} = -\ P_1\ p_1(x_{1o})\ , \tag{4-37}$$

and

$$b_{22} = P_1\ p_1(x_{2o}) + P_2\ p_2(x_{2o})\ . \tag{4-38}$$

In order that $\underline{B}$ be a symmetric matrix,*

$$b_{12} = b_{21} \tag{4-39}$$

$$p_1(x_{1o}) = p_1(x_{2o}) \ , \tag{4-40}$$

which would not generally be satisfied for arbitrary input statistics. With the definition

$$M_i = E[V|\text{"i" transmitted}] \ , \quad i = 1,2,3, \tag{4-41}$$

it appears that Equation (4-40) can be assured only if

$$\text{(i)} \quad p_o(V + M_o) = p_2(V + M_2) \tag{4-42}$$

$$\text{(ii)} \quad M_1 = \frac{M_o + M_2}{2} \tag{4-43}$$

$$\text{(iii)} \quad P_o = P_2 \tag{4-44}$$

---

*Many authors, such as Cramer (26), limit, by definition, the positive definite, positive semi-definite and non-negative definite properties to symmetric matrices only. Using this definition the $\underline{B}$ matrix would be required to be symmetric by Condition 5. Other authors, such as Derusso, Roy and Close (27), do not require that the matrix first be symmetrical so that symmetry of $\underline{B}$ would not be implied by Condition 5. However, the theorem later requires that an orthogonal matrix, $\underline{P}$, be found to diagonalize $\underline{B}$. In order that such an orthogonal matrix exists, $\underline{B}$ must be symmetric. Therefore, regardless of the definition of positive definite used, the matrix $\underline{B}$ must be symmetrical.

(iv) $p_o(V)$, $p_1(V)$ and $p_2(V)$ symmetrical about M_o, M_1 and M_2, respectively.

These conditions are met, for example, when the three received signals are gaussianly distributed, with

(a) equal but unknown spacing,

(b) $P_o = P_2$, but unknown,

(c) $\sigma_o^2 = \sigma_2^2$, but unknown.

Under the four conditions, (i) through (iv),

$$p_1(x_{1o}) = p_1(x_{2o}) \tag{4-45}$$

and

$$p_o(x_{1o}) = p_2(x_{2o}) \tag{4-46}$$

so that

$$b_{11} = b_{22} , \tag{4-47}$$

$$b_{12} = b_{21} , \tag{4-48}$$

and

$$\underline{B} = \begin{bmatrix} P_o\, p_o(x_{1o}) + P_1\, p_1(x_{1o}) & -P_1\, p_1(x_{1o}) \\ -P_1\, p_1(x_{1o}) & P_o\, p_o(x_{1o}) + P_1\, p_1(x_{1o}) \end{bmatrix} \qquad (4\text{-}49)$$

Using the orthogonal matrix, $\underline{P}$, given by Equation (4-22)

$$\underline{P}^{-1}\, \underline{B}\, \underline{P} = \begin{bmatrix} P_o\, p_o(x_{1o}) + 2\, P_1\, p_1(x_{1o}) & 0 \\ 0 & P_o\, p_o(x_{1o}) \end{bmatrix} \qquad (4\text{-}50)$$

so that the characteristic roots of $\underline{B}$ are

$$b_1 = P_o\, p_o(x_{1o}) + 2\, P_1\, p_1(x_{1o}) \qquad (4\text{-}51)$$

and

$$b_2 = P_o\, p_o(x_{1o}) \; . \qquad (4\text{-}52)$$

Since both are positive, $\underline{B}$ is positive definite as required by Condition 5 and

$$b_1 > b_2$$

as required later.

Since $\underline{P}$ given by Equation (4-22) is the matrix which diagonalizes $\underline{B}$, $\underline{\pi}^*$ is given by Equation (4-24), so that

$$\pi^*_{11} = 3\, P_o \int_{x_{1o}}^{\infty} p_o(V)\, dV \ ,$$

$$\pi^*_{12} = \pi^*_{21} = 0 \ ,$$

$$\pi^*_{22} = P_o \int_{x_{1o}}^{\infty} p_o(V)\, dV \ ,$$

and

$$\underline{Q} = \begin{bmatrix} q_{11} & 0 \\ 0 & q_{22} \end{bmatrix}$$

where

$$q_{11} = \frac{A^2\, \pi^*_{11}}{2\, A\, b_1 - 1}$$

$$q_{11} = \frac{3\, A^2\, P_o \int_{x_{1o}}^{\infty} p_o(V)\, dV}{2A[P_o p_o(x_{1o}) + 2P_1 p_1(x_{1o})] - 1} \qquad (4\text{-}53)$$

and

$$q_{22} = \frac{A^2\, \pi^*_{22}}{2\, A\, b_2 - 1}$$

$$q_{22} = \frac{A^2 P_o \int_{x_{1o}}^{\infty} p_o(V)\, dV}{2AP_o p_o(x_{1o}) - 1} \, . \tag{4-54}$$

Then

$$\underline{Q}^* = \underline{P}\, \underline{Q}\, \underline{P}^{-1}$$

$$\underline{Q}^* = \begin{bmatrix} \dfrac{q_{11} + q_{22}}{2} & \dfrac{-q_{11} + q_{22}}{2} \\[2ex] \dfrac{-q_{11} + q_{22}}{2} & \dfrac{q_{11} + q_{22}}{2} \end{bmatrix} \tag{4-55}$$

so that

$$q^*_{11} = q^*_{22} = \frac{3\, A^2\, P_o \int_{x_{10}}^{\infty} p_o(V)\, dV}{4A[P_o p_o(x_{1o}) + 2P_1 p_1(x_{1o})] - 2} \tag{4-56}$$

$$+ \frac{A^2\, P_o \int_{x_{10}}^{\infty} p_o(V)\, dV}{4AP_o p_o(x_{1o}) - 2} \, ,$$

$$q^*_{12} = q^*_{21} = \frac{-\, 3\, A^2\, P_o \int_{x_{10}}^{\infty} p_o(V)\, dV}{4A[P_o\, p_o(x_{1o}) + 2P_1 p_1(x_{1o})] - 2} \tag{4-57}$$

$$+ \frac{A^2 P_o \int_{x_{1o}}^{\infty} p_o(V)\, dV}{4AP_o p_o(x_{1o}) - 2}$$

and

$$\mathrm{Var}\,(x_1 - x_{1o}) = \mathrm{Var}\,(x_2 - x_{2o}) = \frac{q^*_{11}}{n}\,, \qquad (4\text{-}58)$$

and

$$\mathrm{Cov}\,(x_1 - x_{1o})(x_2 - x_{2o}) = \frac{q^*_{12}}{n}\,. \qquad (4\text{-}59)$$

Minimizing the error variance with respect to A results in a complicated expression. However, minimizing each term of Equation (4-56) separately provides a lower bound on the minimum error variance. Thus, for $i = 1,2$,

$$\mathrm{Min\ Var}\,(x_i - x_{io}) > \frac{1}{n}\left[\frac{3\,P_o \int_{x_{1o}}^{\infty} p_o(V)\, dV}{2[P_o p_o(x_{1o}) + P_1 p_1(x_{1o})]^2} + \frac{P_o \int_{x_{1o}}^{\infty} p_o(V)\, dV}{2[P_o p_o(x_{1o})]^2}\right]. \qquad (4\text{-}60)$$

This result is similar to the minimum error variance, given by Equation

(3-98), for the analogous one dimensional criterion.

<u>Criterion: Minimize $\{P_o\, P[\varepsilon|0] + P_1\, P[\varepsilon|1] + P_2\, P[\varepsilon|2]\}$</u>

The two dimensional average probability of error is given by

$$R(\underline{x}) = P_o \int_{x_1}^{\infty} p_o(V)\, dV + P_1 \int_{-\infty}^{x_1} p_1(V)\, dV \tag{4-61}$$

$$+ P_1 \int_{x_2}^{\infty} p_1(V)\, dV + P_2 \int_{-\infty}^{x_2} p_2(V)\, dV \; .$$

The optimum thresholds are $\underline{x}_o$ such that this expression is minimized. A necessary condition is that

$$\left.\frac{\partial R(\underline{x})}{\partial x_1}\right|_{\underline{x}=\underline{x}_o} = 0 \tag{4-62}$$

and

$$\left.\frac{\partial R(\underline{x})}{\partial x_2}\right|_{\underline{x}=\underline{x}_o} = 0 \; , \tag{4-63}$$

if the partial derivatives are defined, so that

$$P_o\, p_o(x_{1o}) = P_1\, p_1(x_{1o}) \tag{4-64}$$

and

$$P_1\, p_1(x_{2o}) = P_2\, p_2(x_{2o}) \quad . \qquad (4\text{-}65)$$

The multidimensional Kiefer-Wolfowitz procedure determines the vector for which a scalar function, $f(\underline{x})$, attains its unique maximum. Since $R(\underline{x})$ is assumed to possess a unique minimum, the procedure can be applied by defining $y(\underline{x})$ such that

$$f(\underline{x}) = E[y(\underline{x})] \ ,$$

$$f(\underline{x}) = -\ R(\underline{x}) \ . \qquad (4\text{-}66)$$

Receiver Structure

Define the random variable

$$y(\underline{x}) = \begin{cases} -\ 1 \text{ if incorrect decision made.} \\ \ \ 0 \text{ if correct decision made.} \end{cases} \qquad (4\text{-}67)$$

so that

$$f(\underline{x}) = E[y(\underline{x})]$$

$$f(\underline{x}) = -\ R(\underline{x}) \ , \qquad (4\text{-}68)$$

and $f(\underline{x})$ has a unique maximum at $\underline{x}_o$. With the sequences $\{a_n\}$ and $\{c_n\}$

properly defined, the thresholds are adjusted according to

$$x_{1,n+1} = x_{1,n} - \frac{a_n}{c_n}\left[y(x_1 - c_n, x_2) - y(x_1 + c_n, x_2)\right] \tag{4-69}$$

and

$$x_{2,n+1} = x_{2,n} - \frac{a_n}{c_n}\left[y(x_1, x_2 - c_n) - y(x_1, x_2 + c_n)\right] . \tag{4-70}$$

With $\underline{x}_1$ an arbitrary vector, the recursive processes converge to $\underline{x}_o$ with probability one as n approaches infinity whenever Conditions 1 and 2 are satisfied. The theorem assumes that each of the four random variables in Equations (4-69) and (4-70) is generated from separate input observations. Thus, the two thresholds are adjusted every four signal intervals.

Equation (A6-7) of Condition 1 requires that the norm of the difference between $f(\underline{x})$ at any two values of $\underline{x}$, such that $|\underline{x}|$ is finite, is bounded. Thus $f(\underline{x})$ may have only finite discontinuities for $|\underline{x}|$ finite but infinite discontinuities for $|\underline{x}|$ infinite. $R(\underline{x})$ has only finite discontinuities, if any, for all values of $\underline{x}$. Thus, this part of the condition is satisfied. To evaluate the requirement of Equation (A6-8), define

$$\underline{M}(\underline{x},a) \triangleq \begin{bmatrix} f(x_1 + a, x_2) \\ f(x_1, x_2 + a) \end{bmatrix} ,$$

and let

$$g \triangleq (\underline{M}(\underline{x},-\varepsilon) - \underline{M}(\underline{x},\varepsilon), \underline{x} - \underline{x}_o) \cdot \tag{4-71}$$

The second part of Condition 1 will be satisfied if g is positive for all $\underline{x}$ such that

$$\varepsilon < \|\underline{x} - \underline{x}_o\| < \infty .$$

Then, from the definition,

$$g = \left(f(x_1-\varepsilon,x_2) - f(x_1+\varepsilon,x_2) \quad f(x_1,x_2-\varepsilon) - f(x_1,x_2+\varepsilon)\right) \begin{bmatrix} x_1 - x_{1o} \\ x_2 - x_{2o} \end{bmatrix}$$

$$g = (x_1-x_{1o}) [f(x_1-\varepsilon,x_2) - f(x_1+\varepsilon,x_2)]$$

$$+ (x_2-x_{2o}) [f(x_1,x_2-\varepsilon) - f(x_1,x_2+\varepsilon)]$$

$$g = (x_1-x_{1o}) [- P_o \int_{x_1-\varepsilon}^{\infty} p_o(V)\, dV - P_1 \int_{-\infty}^{x_1-\varepsilon} p_1(V)\, dV$$

$$+ P_o \int_{x_1+\varepsilon}^{\infty} p_o(V)\, dV + P_1 \int_{-\infty}^{x_1+\varepsilon} p_1(V)\, dV]$$

$$+ (x_2-x_{2o}) [-P_1 \int_{x_2-\varepsilon}^{\infty} p_1(V)\, dV - P_2 \int_{-\infty}^{x_2-\varepsilon} p_2(V)\, dV$$

$$+ P_1 \int_{x_2+\varepsilon}^{\infty} p_1(V)\, dV + P_2 \int_{-\infty}^{x_2+\varepsilon} p_2(V)\, dV]$$

which can be written as

$$g = (x_1 - x_{1o})\, [P_1 \int_{x_1-\varepsilon}^{x_1+\varepsilon} p_1(V)\, dV - P_o \int_{x_1-\varepsilon}^{x_1+\varepsilon} p_o(V)\, dV] \tag{4-72}$$

$$+ (x_2 - x_{2o})\, [P_2 \int_{x_2-\varepsilon}^{x_2+\varepsilon} p_2(V)\, dV - P_1 \int_{x_2-\varepsilon}^{x_2+\varepsilon} p_1(V)\, dV] \,.$$

For ε small, this becomes

$$g = (x_1 - x_{1o})\, 2\varepsilon [P_1\, p_1(x_1) - P_o\, p_o(x_1)] \tag{4-73}$$

$$+ (x_2 - x_{2o})\, 2\varepsilon [P_2\, p_2(x_2) - P_1\, p_1(x_2)]$$

and both terms are positive, and Condition 1 satisfied, if unique solutions exist to

$$P_1\, p_1(x_1) = P_o\, p_o(x_1)$$

$$P_2\, p_2(x_2) = P_1\, p_1(x_2)$$

which is also, from Equations (4-64) and (4-65), the condition necessary to assure a unique solution, $\underline{x}_o$, to the criterion.

Condition 2 is always satisfied because $y(\underline{x})$ takes on only finite values. Therefore, with both conditions satisfied, the thresholds defined by Equations (4-69) and (4-70) converge to $\underline{x}_o$. The receiver block diagram based on these equations is shown in Figure 28.

Convergence

Condition 5 is always satisfied because $y(\underline{x})$ takes on only finite values. To examine Condition 6, let

$$h \triangleq \left(\underline{x} - \underline{x}_o,\ \frac{1}{c}\,(\underline{M}(\underline{x},-c) - \underline{M}(\underline{x},c))\right)$$

so that the condition is satisfied if there exist positive numbers ε, c_o and K_1, such that for

$$c \leq c_o \ ,$$

and

$$c < \|\underline{x} - \underline{x}_o\| < \varepsilon \ ,$$

$$h > K_1\,(x_1 - x_{1o})^2 + K_1\,(x_2 - x_{2o})^2 \ .$$

Then, from the definition,

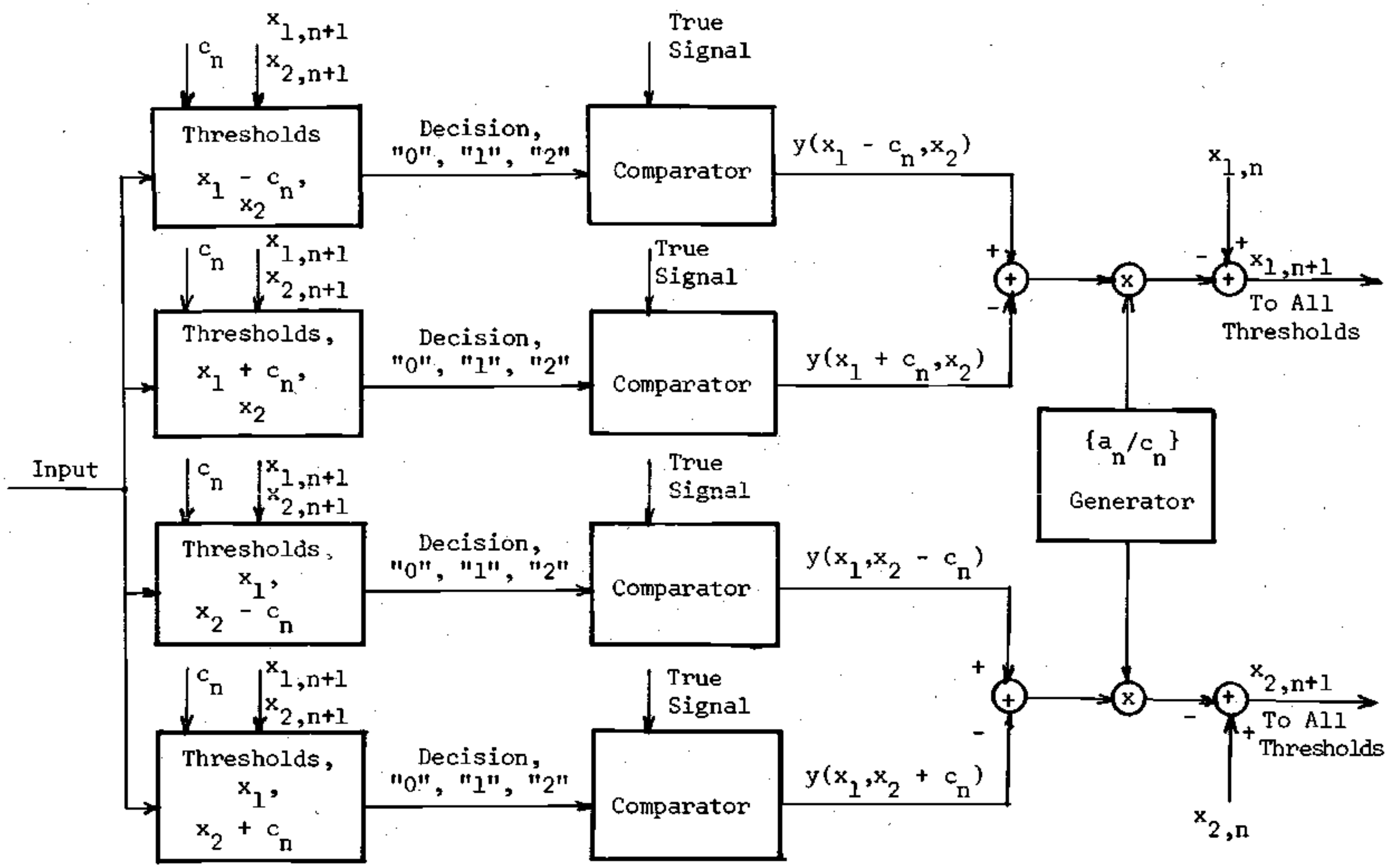


Figure 28. Receiver Based on the Multidimensional Kiefer-Wolfowitz Technique

$$h = \frac{f(x_1-c,x_2) - f(x_1+c,x_2)}{c}(x_1 - x_{1o})$$

$$+ \frac{f(x_1,x_2-c) - f(x_1,x_2+c)}{c}(x_2 - x_{2o}) .$$

Using the representation of $f(\underline{x})$ given by Equation (A6-13) this becomes

$$h = \frac{x_1 - x_{1o}}{c}\{- b_{11}(x_1-c-x_{1o})^2 - b_{22}(x_2-x_{2o})^2$$

$$+ \delta(x_1-c,x_2;\underline{x}_o) + b_{11}(x_1+c-x_{1o})^2$$

$$+ b_{22}(x_2-x_{2o})^2 - \delta(x_1+c,x_2;\underline{x}_o)\}$$

$$+ \frac{x_2 - x_{2o}}{c}\{- b_{11}(x_1-x_{1o})^2 - b_{22}(x_2-c-x_{2o})^2$$

$$+ \delta(x_1,x_2-c;\underline{x}_o) + b_{11}(x_1-x_{1o})^2$$

$$+ b_{22}(x_2+c-x_{2o})^2 - \delta(x_1,x_2+c;\underline{x}_o)\}$$

which can be written as

$$h = 4b_{11}(x_1-x_{1o})^2 + (x_1-x_{1o})\left[\frac{\delta(x_1-c,x_2;\underline{x}_o) - \delta(x_1+c,x_2;\underline{x}_o)}{c}\right]$$

$$+ 4b_{22}(x_2-x_{2o})^2 + (x_2-x_{2o})\left[\frac{\delta(x_1,x_2-c;\underline{x}_o) - \delta(x_1,x_2+c;\underline{x}_o)}{c}\right] .$$

For c sufficiently small,

$$h = 4b_{11}(x_1-x_{1o})^2 + 4b_{22}(x_2-x_{2o})^2 \tag{4-74}$$

$$- 2(x_1-x_{1o}) \frac{\partial\delta(\underline{x},\underline{x}_o)}{\partial x_1} - 2(x_2-x_{2o}) \frac{\partial\delta(\underline{x},\underline{x}_o)}{\partial x_2} .$$

The last two terms can be made arbitrarily small through the choice of ε in Equation (A6-19), so that

$$h > K_1 (x_1-x_{1o})^2 + K_1 (x_2-x_{2o})^2 , \tag{4-75}$$

and Condition 6 is satisfied, if

$$K_1 < 4 \inf \{b_{11}, b_{22}\} . \tag{4-76}$$

Therefore, as long as neither b_{11} nor b_{22} is zero, the condition can be satisfied.

Condition 3 requires that $f(\underline{x})$ can be represented as in Equation (A6-13) where $\underline{B}$ is a positive definite matrix. Thus

$$f(\underline{x}) = \alpha_o - b_{11}(x_1-x_{1o})^2 - (b_{12}+b_{21})(x_1-x_{1o})(x_2-x_{2o}) \tag{4-77}$$

$$- b_{22}(x_2-x_{2o})^2 + \delta(\underline{x},\underline{x}_o) .$$

Taking the second partial derivative with respect to x_1 gives

$$\frac{\partial^2 f(\underline{x})}{\partial\ x_1^{\ 2}} = -\ b_{11} + \frac{\partial^2 \delta(\underline{x},\underline{x}_o)}{\partial\ x_1^{\ 2}} \tag{4-78}$$

and evaluating at $\underline{x}_o$,

$$\left.\frac{\partial^2 f(\underline{x})}{\partial\ x_1^{\ 2}}\right|_{\underline{x}=\underline{x}_o} = -\ b_{11} + \left.\frac{\partial^2 \delta(\underline{x},\underline{x}_o)}{\partial\ x_1^{\ 2}}\right|_{\underline{x}=\underline{x}_o} . \tag{4-79}$$

But the requirement imposed on $\delta(\underline{x},\underline{x}_o)$ by Equation (A6-14) is such that

$$\left.\frac{\partial^2 \delta(\underline{x},\underline{x}_o)}{\partial\ x_1^{\ 2}}\right|_{\underline{x}=\underline{x}_o} = \left.\frac{\partial^2 \delta(\underline{x},\underline{x}_o)}{\partial\ x_1\ \partial x_2}\right|_{\underline{x}=\underline{x}_o} = \left.\frac{\partial^2 \delta(\underline{x},\underline{x}_o)}{\partial\ x_2^{\ 2}}\right|_{\underline{x}=\underline{x}_o} = 0\ , \tag{4-80}$$

so that

$$b_{11} = -\left.\frac{\partial^2 f(\underline{x})}{\partial\ x_1^{\ 2}}\right|_{\underline{x}=\underline{x}_o} , \tag{4-81}$$

and similarly,

$$b_{12} + b_{21} = -\left.\frac{\partial^2 f(\underline{x})}{\partial\ x_1\ \partial\ x_2}\right|_{\underline{x}=\underline{x}_o} \tag{4-82}$$

and

$$b_{22} = - \left. \frac{\partial^2 f(\underline{x})}{\partial x_2^2} \right|_{\underline{x}=\underline{x}_o} . \tag{4-83}$$

The requirement that $\underline{B}$ be symmetric in this case presents no problem, since only the sum of b_{12} and b_{21} is defined. Thus, to make $\underline{B}$ symmetric, let

$$b_{12} = b_{21} = - \frac{1}{2} \left. \frac{\partial^2 f(\underline{x})}{\partial x_1 \, \partial x_2} \right|_{\underline{x}=\underline{x}_o} . \tag{4-84}$$

These elements are then determined from the definition of $f(\underline{x})$, Equations (4-66) and (4-61), with the results

$$b_{11} = P_1 \, p_1'(x_{1o}) - P_o \, p_o'(x_{1o}) , \tag{4-85}$$

$$b_{12} = b_{21} = 0 , \tag{4-86}$$

and

$$b_{22} = P_2 \, p_2'(x_{2o}) - P_1 \, p_1'(x_{2o}) . \tag{4-87}$$

Since $\underline{B}$ is already a diagonal matrix, b_{11} and b_{22} are its characteristic roots, which must be positive for $\underline{B}$ to be positive definite. For unimodal density functions and reasonably well separated signal emplitudes, the optimum thresholds lie between the peaks and

$$p_1'(x_{1o}) > 0 \ , \tag{4-88}$$

$$p_o'(x_{1o}) < 0 \ , \tag{4-89}$$

so that

$$b_{11} > 0 \tag{4-90}$$

and

$$p_2'(x_{2o}) > 0 \tag{4-91}$$

$$p_1'(x_{2o}) < 0 \tag{4-92}$$

so that

$$b_{22} > 0 \ . \tag{4-93}$$

In addition, since $\underline{B}$ is diagonal, the required orthogonal matrix is

$$\underline{P} = \underline{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} . \tag{4-94}$$

Using the definition of $\underline{Z}(\underline{x},a)$ given by Equation (A6-11) and Condition 4, the π_{11} element of the $\underline{\pi}$ matrix is defined as

$$\pi_{11} = \lim_{\substack{\underline{x}\to\underline{x}_o \\ c\to 0}} E\{[y(x_1-c,x_2) - f(x_1-c,x_2) - y(x_1+c,x_2) + f(x_1+c,x_2)]^2\} \ . \tag{4-95}$$

But

$$\lim_{c\to 0} f(x_1+c,x_2) = \lim_{c\to 0} f(x_1-c,x_2) \tag{4-96}$$

so that

$$\pi_{11} = \lim_{\substack{\underline{x}\to\underline{x}_o \\ c\to 0}} E\{[y(x_1-c,x_2) - y(x_1+c,x_2)]^2\} \ . \tag{4-97}$$

The difference of the two random variable terms does not vanish in the limit because the terms are generated from independent observations. Thus

$$\pi_{11} = \lim_{\substack{\underline{x}\to\underline{x}_o \\ c\to 0}} \{- f(x_1-c,x_2) - 2f(x_1-c,x_2)\, f(x_1+c,x_2) - f(x_1+c,x_2)\}$$

$$\pi_{11} = -f(\underline{x}_o) - 2f^2(\underline{x}_o) - f(\underline{x}_o)$$

$$\pi_{11} = 2[R(\underline{x}_o) - R^2(\underline{x}_o)] \ . \tag{4-98}$$

Similarly

$$\pi_{22} = 2[R(\underline{x}_o) - R^2(\underline{x}_o)] \tag{4-99}$$

and

$$\pi_{12} = \lim_{\substack{\underline{x}\to\underline{x}_o \\ c\to 0}} E\{[y(x_1-c,x_2) - y(x_1+c,x_2)]x$$

$$[y(x_1,x_2-c) - y(x_1,x_2+c)]\}$$

$$+ \begin{bmatrix}\text{Terms which become} \\ \text{zero in the limit.}\end{bmatrix}$$

so that

$$\pi_{12} = \lim_{\substack{\underline{x}\to\underline{x}_o \\ c\to 0}} \{f(x_1-c,x_2)\ f(x_1,x_2-c) - f(x_1+c,x_2)\ f(x_1,x_2-c)$$

$$- f(x_1-c,x_2)\ f(x_1,x_2-c) + f(x_1+c,x_2)\ f(x_1,x_2+c)\}$$

$$\pi_{12} = f^2(\underline{x}_o) - f^2(\underline{x}_o) - f^2(\underline{x}_o) + f^2(\underline{x}_o)$$

$$\pi_{12} = 0 \tag{4-100}$$

and similarly

$$\pi_{21} = 0\ . \tag{4-101}$$

Thus

$$\underline{\pi} = \begin{bmatrix} 2[R(\underline{x}_o) - R^2(\underline{x}_o)] & 0 \\ 0 & 2[R(\underline{x}_o) - R^2(\underline{x}_o)] \end{bmatrix} . \quad (4\text{-}102)$$

Since the characteristic roots, π_{11} and π_{22}, are positive, $\underline{\pi}$ is positive definite and therefore satisfies the requirement that it be non-negative definite. Since $\underline{P}$ is the identity matrix,

$$\underline{\pi}^* = \underline{P}^{-1}\, \underline{\pi}\, \underline{P} = \underline{\pi} \quad (4\text{-}103)$$

and

$$\underline{W}^* = \underline{P}^{-1}\, \underline{W}\, \underline{P} = \underline{W} \,. \quad (4\text{-}104)$$

Let

$$\lambda_1 = \sup\, \{b_{11}, b_{22}\} \,, \quad (4\text{-}105)$$

$$\lambda_2 = \inf\, \{b_{11}, b_{22}\} \,, \quad (4\text{-}106)$$

$$a_n = A/n \,,$$

then

$$w^*_{11} = \frac{2\ A^2[R(\underline{x}_o) - R^2(\underline{x}_o)]}{8\ A\ \lambda_1 - 1}\ , \tag{4-107}$$

$$w^*_{22} = \frac{2\ A^2[R(\underline{x}_o) - R^2(\underline{x}_o)]}{8\ A\ \lambda_2 - 1} \tag{4-108}$$

and

$$w^*_{12} = w^*_{21} = 0\ , \tag{4-109}$$

so that

$$\mathrm{Var}\ (x_1 - x_{1o}) = \frac{1}{n^{\frac{1}{2}}}\ \frac{2\ A^2[R(\underline{x}_o) - R^2(\underline{x}_o)]}{8\ A\ \lambda_1 - 1}\ , \tag{4-110}$$

$$\mathrm{Var}\ (x_2 - x_{2o}) = \frac{1}{n^{\frac{1}{2}}}\ \frac{2\ A^2[R(\underline{x}_o) - R^2(\underline{x}_o)]}{8\ A\ \lambda_2 - 1}\ , \tag{4-111}$$

and

$$\mathrm{Cov}\ (x_1 - x_{1o})\ (x_2 - x_{2o}) = 0\ . \tag{4-112}$$

Equations (4-110) and (4-111) are similar to the resulting error variance, Equation (3-183), for the analogous one dimensional criterion.

For the special case where the input density functions are equal and equally spaced gaussian densities and the three signals are equally likely to occur.

$$P_1\, p_1'(x_{1o}) = P_2\, p_2'(x_{2o}) = -\, P_o\, p_o'(x_{1o}) = -\, P_1\, p_1'(x_{2o})$$

so that

$$b_{11} = b_{22} = 2\, P_1\, p_1'(x_{1o}) \; , \tag{4-113}$$

$$\mathrm{Var}\,(x_1 - x_{1o}) = \mathrm{Var}\,(x_2 - x_{2o}) \; ,$$

and

$$\mathrm{Var}\,(x_i - x_{io}) = \frac{1}{\frac{1}{n^2}} \frac{2\, A^2 [R(\underline{x}_o) - R^2(\underline{x}_o)]}{16\, A\, P_1\, p_1'(x_{1o}) - 1} \; , \tag{4-114}$$

for $i = 1,2$. If A is selected such that

$$A = 1/4[2\, P_1\, p_1'(x_{10})] \; ,$$

for $i = 1,2$,

$$\mathrm{Min\ Var}\,(x_i - x_{io}) = \frac{1}{\frac{1}{n^2}} \frac{R(\underline{x}_o) - R^2(\underline{x}_o)}{32[P_1\, p_1'(x_{1o})]^2} \; , \tag{4-115}$$

which has the same form as Equation (3-184), when evaluated for equivalent input conditions.

CHAPTER V

DISCUSSION AND RECOMMENDATIONS

The mathematical theorems of stochastic approximation have been shown to be readily applicable to the single threshold, adaptive receiver problem for many criteria of engineering importance. Although the theorems themselves are somewhat abstract and complex, the iterative form of the resulting receivers make their synthesis and operation simple. The only storage required is a one interval delay of the threshold setting; calculations involve only simple operations of simple quantities. This is in contrast to most pattern recognition techniques which are simple and straightforward in principle, but result in receivers requiring extensive calculations and storage.

For the one dimensional problem, the critical step in applying the theorems is in defining a random variable, $T(x)$, whose expectation is properly related to the performance criterion of interest. With the theorems providing only conditions on $T(x)$ and no guidance for selecting it, this usually involves a trial and error process. Once this random variable has been defined, synthesis and evaluation of the receiver is straightforward.

In the heuristic development of stochastic approximation in Chapter II, $T(x)$ was considered a "noisy" estimate of $N(x)$. In applications, the distinction between $T(x)$ and the noise corrupted input signals must be kept clear. The input signal is composed of a transmitted signal which is corrupted in some random manner; thus it is physically a noisy signal. This random signal is compared to a threshold and a decision made regarding the symbol transmitted. On the other hand, $T(x)$ is a random variable internally generated on the basis of the decision and the true signal; thus it is physically a "clean" signal. Still it is convenient to think of $T(x)$ as a "noisy" representation of $N(x)$ since it is a random variable whose expectation is $N(x)$.

Multidimensional problems likewise require proper generation of a random variable, although in some cases it may be vector valued. The resulting receivers are only slightly more complicated than the one dimensional receivers and are still considerably simpler than those resulting from other techniques. However, interpretation of the mathematical requirements and their consequences is not so straightforward. In fact, the functional form of some of the resulting equations, and the extremely non-linear manner in which the thresholds enter into them, makes explicit determination of requirements on the input statistics difficult. It may be that the multidimensional theorems are applicable to as wide a class of input statistics as are the one dimensional theorems. For the criteria considered, only sufficient, and not necessary, conditions on the input statistics were obtained. The class of input signals which satisfies the sufficient conditions is relatively small, but important in engineering applications.

Obtaining the Training Sequence

During training, both the actual received signal and the true transmitted sequence must be available at the receiver. Of course, all of the input statistics which enter into the performance criterion must be the same during training as they are during operation. Since the need for an adaptive receiver arises only when complete statistical information is not available, artificial training signals cannot generally be properly generated. Training must therefore normally be accomplished in the operating environment.

One method of obtaining the true sequence would be to switch the transmitter to a special error correcting code mode during training. A second receiver, with decoding and error correcting capability would provide an essentially error-free sequence to the receiver being trained. Occasional errors in the sequence would have little effect on the convergence rate since threshold adjustments are often in the wrong direction, even with an accurate sequence.

A second method would be to slow down the transmission data rate during training. Again a second receiver channel, including a narrow-band filter, would be required. With the reduced data rate, the narrow-band filter reduces the effects of noise without affecting the signal amplitude. A threshold type decision device, operating with a reduced probability of error, would then provide a sufficiently accurate sequence to the adaptive receiver.

Some criteria are such that no special training techniques are required. For example, a radar system, whose performance criterion involves only noise statistics, may be trained any time there are no, or few, targets within range. In fact, for many systems, the effective input noise is primarily due to internally generated, front-end noise.

In such cases, training may even be accomplished in a RF controlled environment, without transmitter or antenna.

Limitations

Stationary Input Statistics

The most severe limitation of the stochastic approximation technique stems from the requirement that the input statistics be stationary. Although the requirement only concerns the statistics during training, the threshold setting obtained is optimum or near optimum only for as long as the statistics remain essentially unchanged. In practical applications the statistics will always change with time to some extent. The receiver must be periodically retrained to maintain satisfactory performance. This shortcoming is characteristic not only of the stochastic approximation technique, but on any supervised learning technique.

Several factors influence the frequency of training. The amount of training time required to set the threshold within acceptable limits is of prime importance. Since the error variance is a function of the number of iterations in the training period, the system's data rate will determine the actual amount of time required. Thus, a system with a high data rate will require a smaller percentage of available operating time for training than a system with a low data rate. The rate at which the input statistics change and the sensitivity of receiver performance to threshold accuracy will also influence the required training frequency.

For systems in which the statistics change slowly compared to the required training time, the percentage of available operating time devoted to training is low and a supervised learning technique is practical. For systems in which the statistics change at a rate comparable to the required training time, training would be almost continuous and such techniques are not practical.

In some applications, changes occur from signal interval to signal interval in a random manner, which can be described by a probability density function. For such cases, the randomness of these rapid changes is combined with the randomness of the transmitted signal and channel noise to determine a complete a posteriori input probability density function for each symbol. The receiver converges to the optimum threshold in the normal manner; the only difference is that the value of the optimum setting is influenced by the additional random factor.

Independent Observations

The finite bandwidths of real systems make completely independent observations impossible to obtain in practice. On the other hand, the amount of dependence generally decreases rapidly with an increase in time between observations. The dependence can therefore be made arbitrarily small so that this requirement is not a severe limitation.

This requirement has been reduced somewhat by Sakrison (28). The weaker restriction cannot be expressed directly in terms of the observations or their distributions but involves a measure of the predictability of the observations. Yet he does note that most real processes would be expected to satisfy the requirement.

Recommendations

There appear to be several areas deserving further research, both mathematical and engineering. Sacks' asymptotic distribution theorems hold only for A greater than some minimum, non-zero value, whereas the basic theorems assure convergence for all positive values of A. Convergence properties for the smaller values of A should be determined.

Sakrison showed that the processes will converge even if there is some dependence between observations. The effect of dependent observations on convergence rate should be evaluated, either through mathematical derivation or system simulation.

Increased interest in pattern recognition makes the multidimensional processes the most promising area of future study. Mathematical work should be directed towards reducing the conditions of both the basic theorems and the asymptotic distribution theorems, so that the theorems can be applied to a larger class of input signals. In addition, the conditions should be determined explicitly in terms of the input statistics to enable more direct application of the theorems.

Computer simulation or actual system studies should be carried out, prior to or in conjunction with the theoretical work. Results would likely suggest areas in which further theoretical work should be concentrated and would provide greater engineering insight into the multidimensional processes.

APPENDIX

APPENDIX I

ROBBINS AND MONRO'S FIRST THEOREM

Let $N(x)$ be a fixed but unknown real-valued function of x and x_o be an unknown value of x such that, for some

$$\delta > 0 ,$$

$$N(x) \leq -\delta \quad \text{for finite } x < x_o \tag{A1-1}$$

and

$$N(x) \geq \delta \quad \text{for finite } x > x_o . \tag{A1-2}$$

Define a random variable, $T(x)$, such that

$$E[T(x)] = N(x) \tag{A1-3}$$

for all x, and for some finite positive constant, C,

$$\Pr[|T(x)| \leq C] = 1 \tag{A1-4}$$

for all x. Define a sequence, $\{a_n\}$, such that

$$\sum_{n=1}^{\infty} a_n = \infty , \tag{A1-5}$$

$$\sum_{n=1}^{\infty} a_n^{\,2} < \infty \; . \tag{A1-6}$$

Then the nonstationary Markov chain defined by the recursive equation

$$x_{n+1} = x_n - a_n \; T(x) \; , \tag{A1-7}$$

where x_1 is an arbitrary, finite number, converges to x_o in mean square and with probability one as n approaches infinity.

Robbins and Monro's second theorem, stated in Chapter II, requires, by Equations (2-8), (2-9) and (2-10), that N(x) be non-decreasing and that $N(x_o)$ and $N'(x_o)$ be defined. Their first theorem requires, by Equations (A1-1) and (A1-2), only that N(x) be positive for finite values of x greater than x_o and negative for finite values of x less than x_o. Thus $N(x_o)$ and $N'(x_o)$ need not be defined and N(x) may asymptotically approach zero as x approaches plus and minus infinity.

APPENDIX II

ASYMPTOTIC BEHAVIOR OF MINIMUM ERROR VARIANCE FOR THE CRITERION: MINIMIZE $\{P_o\ P[1|0] + P_1\ P[0|1]\}$

When the input signals have the gaussian density functions given by Equations (3-17) and (3-104),

$$\alpha = \frac{P_o}{\sqrt{2\pi}} \int_{u_o}^{\infty} \exp\ \{-t^2/2\}\ dt \tag{A2-1}$$

$$\beta = \frac{P_1}{\sqrt{2\pi}} \int_{-\infty}^{u_o} \exp\ \{-\frac{(t-M_1/\sigma)^2}{2}\}\ dt \tag{A2-2}$$

For the minimum probability of error criterion, the normalized optimum threshold is determined from

$$P_o\ p_o(u_o) = P_1\ p_1(u_o)\ . \tag{A2-3}$$

The left-hand side is

$$P_o\ p_o(u_o) = P_o\ \frac{1}{\sqrt{2\pi}\ \sigma} \exp\ \{-u_o^2/2\}$$

or, by taking the natural logarithm and rearranging,

$$u_o^2/2 = -\ln[\sqrt{2\pi}\,\sigma\, p_o(u_o)] \tag{A2-4}$$

and

$$u_o = \sqrt{2\left|\ln[\sqrt{2\pi}\,\sigma\, p_o(u_o)]\right|}\ . \tag{A2-5}$$

Rewriting Equation (A2-3)

$$(P_o/P_1)\, p_o(u_o) = p_1(u_o)$$

so that

$$(P_o/P_1)\, p_o(u_o) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(u_o - M_1/\sigma)^2}{2}\right\}$$

or, by taking the natural logarithm and rearranging,

$$\frac{(u_o - M_1/\sigma)^2}{2} = -\ln[(P_o/P_1)\sqrt{2\pi}\,\sigma\, p_o(u_o)] \tag{A2-6}$$

and

$$M_1/\sigma - u_o = \sqrt{2\left|\ln[(P_o/P_1)\sqrt{2\pi}\,\sigma\, p_o(u_o)]\right|}\ . \tag{A2-7}$$

Wozencraft and Jacobs (29) show that the integral

$$q(\rho) \triangleq \int_{\rho}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\{-t^2/2\}\, dt$$

is bounded by

$$\frac{1}{\sqrt{2\pi}\,\rho} \exp\{-\rho^2/2\}\,(1-1/\rho^2) < q(\rho) < \frac{1}{\sqrt{2\pi}\,\sigma} \exp\{-\rho^2/2\} \tag{A2-8}$$

for ρ greater than zero. As ρ becomes large, the bounds become tight and

$$q(\rho) \cong \frac{1}{\sqrt{2\pi}\,\rho} \exp\{-\rho^2/2\} . \tag{A2-9}$$

Therefore, for large u_o (M_1 large),

$$\alpha \cong \frac{P_o}{\sqrt{2\pi}} \frac{1}{u_o} \exp\{-u_o^2/2\}$$

or, using Equations (A2-4) and (A2-5),

$$\alpha \cong \frac{P_o}{\sqrt{2\pi}} \frac{\exp\{\ln \sqrt{2\pi}\,\sigma\, p_o(u_o)\}}{\sqrt{2\left|\ln[\sqrt{2\pi}\,\sigma\, p_o(u_o)]\right|}}$$

but

$$\exp\{\ln A\} = A ,$$

so that

$$\alpha \cong \frac{\sigma\, P_o\, p_o(u_o)}{\sqrt{2\left|\ln[\sqrt{2\pi}\,\sigma\, p_o(u_o)]\right|}} .$$

For large u_o, $p_o(u_o)$ is very small and

$$|\ln\sqrt{2\pi}\,\sigma| << |\ln p_o(u_o)|$$

so that

$$\alpha \cong \frac{\sigma\, P_o\, p_o(u_o)}{\sqrt{2|\ln p_o(u_o)|}} \quad . \tag{A2-10}$$

β, from Equation (A2-2), may be written as

$$\beta = \frac{P_1}{\sqrt{2\pi}} \int_{\frac{M_1}{\sigma} - u_o}^{\infty} \exp\{-t^2/2\}\, dt$$

so that for large M_1,

$$\beta \cong \frac{P_1}{\sqrt{2\pi}} \frac{1}{(M_1/\sigma - u_o)} \exp\{-\frac{(u_o - M_1/\sigma)^2}{2}\}$$

or, using Equations (A2-6) and (A2-7),

$$\beta \cong \frac{P_1}{\sqrt{2\pi}} \frac{(P_o/P_1)\sqrt{2\pi}\,\sigma\, p_o(u_o)}{\sqrt{2|\ln[(P_o/P_1)\sqrt{2\pi}\,\sigma\, p_o(u_o)]|}} \quad .$$

Noting that for large u_o,

$$\left|\ln[(P_o/P_1)\sqrt{2\pi}\,\sigma\, p_o(u_o)]\right| \cong \left|\ln p_o(u_o)\right| ,$$

this becomes

$$\beta \cong \frac{\sigma\, P_o\, p_o(u_o)}{\sqrt{2|\ln p_o(u_o)|}} , \qquad \text{(A2-11)}$$

which is the same as the expression for α given by Equation (A1-10). Thus for large M_1, α and β are approximately equal and

$$R = \alpha + \beta = 2\alpha = 2\beta \qquad \text{(A2-12)}$$

so that when two observations are used to generate $y(x + c)$ and $y(x - c)$, Equation (3-164), the minimum error variance for large M_1 is

$$\text{Min Var } (x_n - x_o') = \frac{\sigma^4}{nc^2}\frac{\alpha}{D} , \qquad \text{(A2-13)}$$

where

$$D = [P_1\, f_1'(u_o) - P_o\, f_o'(u_o)]^2 . \qquad \text{(A2-14)}$$

Using the approximation of Equation (A2-9) this is

$$\text{Min Var } (x_n - x_o') \cong \frac{\sigma^4}{nc^2}\,\frac{P_o \exp\{-u_o^2/2\}}{\sqrt{2\pi}\, u_o\, D} . \qquad \text{(A2-15)}$$

Using the definition of $f_o(u_o)$ given by Equation (3-160), the minimum error variance when one observation is used, Equation (3-163), becomes

$$\text{Min Var }(x_n - x_o') \cong \frac{\sigma^3}{nc} \frac{P_o \exp\{-u_o^2/2\}}{\sqrt{2\pi}\, D} . \qquad \text{(A2-16)}$$

Therefore the normalized minimum error variance increases slower with M_1 when two observations are used than when one observation is used due to the extra factor of u_o in the denominator. However the actual required training time may or may not be less because of the extra σ/c factor in the complete expression.

APPENDIX III

THE KIEFER-WOLFOWITZ THEOREM

Let M(x) be a fixed but unknown Borel-measureable function with a unique maximum at $x = x_o$ and for $0 < t_o < t_1 < t_2 < \infty$,

$$\inf_{\substack{t_1 \le |x-x_o| \le t_2 \\ 0 < \varepsilon \le t_o}} \frac{(x-x_o)[M(x-\varepsilon) - M(x+\varepsilon)]}{\varepsilon} > 0 . \tag{A3-1}$$

In addition, for all x and suitable D_1 and D_2

$$|M(x+1) - M(x)| < D_1 + D_2\, |x| . \tag{A3-2}$$

Define a random variable, Y(x), such that

$$E[Y(x)] = M(x) \tag{A3-3}$$

for all x, and

$$\sup_x E[Z^2(x)] < \infty , \tag{A3-4}$$

where

$$Z(x) = Y(x) - M(x) . \tag{A3-5}$$

Define two sequences of positive numbers, $\{a_n\}$ and $\{c_n\}$, such that

$$\sum_{n=1}^{\infty} a_n = \infty , \tag{A3-6}$$

$$\lim_{n\to\infty} c_n = 0 \tag{A3-7}$$

$$\sum_{n=1}^{\infty} \left(\frac{a_n}{c_n}\right)^2 < \infty . \tag{A3-8}$$

Then the recursive equation

$$x_{n+1} = x_n - \frac{a_n}{c_n} [Y(x-c_n) - Y(x+c_n)] , \tag{A3-9}$$

with x_1 an arbitrary finite number, converges to x_o with probability one as n approaches infinity.

Kiefer and Wolfowitz (19) introduced this method and proved convergence in probability, under similar conditions. Blum (13) later showed convergence with probability one, under more general conditions. This statement of the theorem is due to Sacks and is a part of his second theorem (21).

APPENDIX IV

CONVERGENCE OF THE KIEFER-WOLFOWITZ PROCEDURE

Condition 1; $M(x)$ has a continuous second derivative in a neighborhood of x_o with

$$M''(x_o) = -2\rho \ . \tag{A4-1}$$

Condition 2; Let

$$Z(x) = Y(x) - M(x) \ , \tag{A4-2}$$

then

$$\lim_{\substack{x \to x_o \\ a \to 0}} E\{[Z(x-a) - Z(x+a)]^2\} = \Delta^2 \tag{A4-3}$$

or, in case $Z(x-c_n)$ and $Z(x+c_n)$ are uncorrelated,

$$\lim_{x \to x_o} E[Z^2(x)] = \frac{\Delta^2}{2} \ . \tag{A4-4}$$

Condition 3;

$$\lim_{R\to\infty} \sup_{k} \int_{\{|Z(x_k)| > R\}} Z^2(x_k)\, dP = 0 \ . \tag{A4-5}$$

If the three conditions are satisfied, the Kiefer-Wolfowitz procedure defined in Appendix III, with

$$c_n = n^{-\frac{1}{4}} \ , \tag{A4-6}$$

and

$$a_n = A/n \ , \tag{A4-7}$$

where A is such that

$$A\,\rho > 1/8 \ , \tag{A4-8}$$

converges to x_o in mean square and with probability one as n becomes infinite and $(x_n - x_o)$ is asymptotically normally distributed with mean zero and

$$\mathrm{Var}\ (x_n - x_o) = \frac{1}{n^{1/2}} \frac{\Delta^2\, A^2}{8\rho\, A - 1} \ . \tag{A4-9}$$

This theorem is due to Sacks and it, together with the theorem of Appendix III, is his theorem number two (21).

The value of A that minimizes this expression is

$$A = 1/4\rho = -1/2M''(x_\Theta) \tag{A4-10}$$

and the minimum value of the error variance is

$$\text{Min Var } (x_n - x_\Theta) = \frac{1}{n^{1/2}} \frac{\Delta^2}{4[M''(x_\Theta)]^2} . \tag{A4-11}$$

APPENDIX V

THE MULTIDIMENSIONAL ROBBINS-MONRO PROCEDURE

Let $\underline{N}(\underline{x})$ be a fixed but unknown vector-valued function of the vector $\underline{x}$ and $\underline{x}_o$ be an unknown value of $\underline{x}$ such that

$$\underline{N}(\underline{x}_o) = \underline{0} . \tag{A5-1}$$

Define a random variable, $\underline{T}(\underline{x})$, such that

$$E[\underline{T}(\underline{x})] = \underline{N}(\underline{x}) \quad \text{for all } \underline{x} \tag{A5-2}$$

and a sequence of positive, real numbers such that

$$\sum_{n=1}^{\infty} a_n = \infty , \tag{A5-3}$$

$$\sum_{n=1}^{\infty} a_n^{\,2} < \infty . \tag{A5-4}$$

Then the recursive equation

$$\underline{x}_{n+1} = \underline{x}_n - a_n \, \underline{T}(\underline{x}) \tag{A5-5}$$

where $\underline{x}_1$ is an arbitrary vector, converges to $\underline{x}_o$ with probability one as n approaches infinity, whenever Conditions 1, 2, and 3 are satisfied. Condition 1: $\underline{N}(\underline{x})$ is Borel-measurable and for every positive ε

$$\inf_{\frac{1}{\varepsilon} > \|\underline{x}-\underline{x}_o\| > \varepsilon} (\underline{x}-\underline{x}_o, \underline{N}(\underline{x})) > 0 \; . \tag{A5-6}$$

Condition 2: There exists a positive constant K_1 such that, for all $\underline{x}$,

$$\|\underline{N}(\underline{x})\| \leq K_1 \|\underline{x}-\underline{x}_o\| \; . \tag{A5-7}$$

Condition 3: Define

$$\underline{Z}(\underline{x}) = \underline{T}(\underline{x}) - \underline{N}(\underline{x}) \; , \tag{A5-8}$$

then

$$\sup_{\underline{x}} E\|\underline{Z}(\underline{x})\|^2 < \infty \; . \tag{A5-9}$$

Condition 4:

$$\lim_{\underline{x}\to\underline{x}_o} E(\underline{Z}(\underline{x}) \; \underline{Z}^T(\underline{x})) = \underline{\pi} \; , \tag{A5-10}$$

where $\underline{\pi}$ is a non-negative definite matrix.

Condition 5: For all $\underline{x}$

$$\underline{N}(\underline{x}) = \underline{B}(\underline{x}-\underline{x}_o) + \underline{\delta}(\underline{x},\underline{x}_o) \qquad \text{(A5-11)}$$

where $\underline{B}$ is a positive definite matrix and

$$\|\underline{\delta}(\underline{x},\underline{x}_o)\| = o(\|\underline{x}-\underline{x}_o\|) \quad \text{as} \quad \|\underline{x}-\underline{x}_o\| \to 0 \ . \qquad \text{(A5-12)}$$

Condition 6:

$$\lim_{R\to\infty} \sup_{k} \int_{\{\|\underline{Z}(\underline{x}_k)\| > R\}} \|\underline{Z}(\underline{x}_k)\|^2 \, dP = 0 \ . \qquad \text{(A5-13)}$$

Let $b_1, b_2, \ldots, b_q$ denote the characteristic roots of $\underline{B}$ in decreasing order. Find an orthogonal matrix $\underline{P}$ such that $\underline{P}^{-1}\,\underline{B}\,\underline{P}$ is a diagonal matrix and let π^*_{ij} be the elements of

$$\underline{\pi}^* = \underline{P}^{-1}\,\underline{\pi}\,\underline{P} \ . \qquad \text{(A5-14)}$$

Define $\underline{Q}$ as the matrix whose (i,j)th element is

$$q_{ij} = \frac{A^2\,\pi^*_{ij}}{Ab_i + Ab_j - 1} \qquad \text{(A5-15)}$$

and

$$\underline{Q}^* = \underline{P}\,\underline{Q}\,\underline{P}^{-1} . \qquad \text{(A5-16)}$$

If the six Conditions are satisfied and

$$a_n = A/n , \qquad \text{(A5-17)}$$

where A is such that

$$Ab_q > \frac{1}{2} , \qquad \text{(A5-18)}$$

$n^{\frac{1}{2}}(\underline{x}_n - \underline{x}_o)$ is asymptotically normally distributed with mean zero and covariance matrix $\underline{Q}^*$.

Blum (22) extended the Robbins-Monro method to the multidimensional case, and proved convergence with probability one. Sacks later determined the asymptotic distribution of the error and it is his fifth theorem (24) stated here.

APPENDIX VI

THE MULTIDIMENSIONAL KIEFER-WOLFOWITZ PROCEDURE

Let $f(\underline{x})$ be a fixed but unknown real valued function of the vector $\underline{x}$ and $\underline{x}_o$ be a value of $\underline{x}$ for which $f(\underline{x})$ has a unique maximum. Define a real valued random variable, $y(\underline{x})$, such that

$$E[y(\underline{x})] = f(\underline{x}) \tag{A6-1}$$

and two sequences of positive, real numbers, $\{a_n\}$ and $\{c_n\}$, satisfying

$$\sum_{n=1}^{\infty} a_n = \infty \tag{A6-2}$$

$$\lim_{n\to\infty} c_n = 0 \tag{A6-3}$$

$$\sum_{n=1}^{\infty} \left(\frac{a_n}{c_n}\right)^2 < \infty . \tag{A6-4}$$

Let $\underline{e}_i$ be the vector whose ith coordinate is 1 and whose other coordinates are 0 and define

$$\underline{Y}(\underline{x},a) = \begin{bmatrix} y(\underline{x} + a\underline{e}_1) \\ y(\underline{x} + a\underline{e}_2) \\ \cdot \\ \cdot \\ \cdot \\ y(\underline{x} + a\underline{e}_q) \end{bmatrix} \tag{A6-5}$$

where q is the dimensionality of $\underline{x}$. Then the recursive equation

$$\underline{x}_{n+1} = \underline{x}_n - \frac{a_n}{c_n}[\underline{Y}(\underline{x},-c_n) - \underline{Y}(\underline{x},c_n)] \tag{A6-6}$$

where $\underline{x}_1$ is an arbitrary vector, and each component of $\underline{Y}(\underline{x},-c_n)$ and $\underline{Y}(\underline{x},c_n)$ is generated from an independent observation, converges to $\underline{x}_o$ with probability one as n approaches infinity, whenever Conditions 1 and 2 are satisfied.

Condition 1: $f(\underline{x})$ is Borel-measurable, has a unique maximum at $\underline{x}_o$, and for some positive constants D_1 and D_2,

$$\|f(\underline{x}+\underline{1}) - f(\underline{x})\| \le D_1 + D_2\|\underline{x}\| \tag{A6-7}$$

and for $0 < \varepsilon_0 < \varepsilon_1 < \varepsilon_2 < \infty$

$$\inf_{\substack{\varepsilon_1 \le \|\underline{x}-\underline{x}_o\| \le \varepsilon_2 \\ 0 < \varepsilon \le \varepsilon_o}} \frac{(\underline{M}(\underline{x},-\varepsilon) - \underline{M}(\underline{x},\varepsilon),\ \underline{x} - \underline{x}_o)}{\varepsilon} > 0 \tag{A6-8}$$

where

$$\underline{M}(\underline{x},a) = \begin{bmatrix} f(\underline{x} + a\underline{e}_1) \\ f(\underline{x} + a\underline{e}_2) \\ \cdot \\ \cdot \\ \cdot \\ f(\underline{x} + a\underline{e}_q) \end{bmatrix} . \tag{A6-9}$$

Condition 2: Let

$$z(\underline{x}) = y(\underline{x}) - f(\underline{x}) \tag{A6-10}$$

and define

$$\underline{Z}(\underline{x},a) = \begin{bmatrix} z(\underline{x} + a\underline{e}_1) \\ z(\underline{x} + a\underline{e}_2) \\ \cdot \\ \cdot \\ \cdot \\ z(\underline{x} + a\underline{e}_q) \end{bmatrix} \tag{A6-11}$$

then

$$\sup_{\underline{x}} E\{\|\underline{Z}(\underline{x},0)\|^2\} < \infty . \tag{A6-12}$$

Condition 3: For all $\underline{x}$

$$f(\underline{x}) = \alpha_o - (\underline{B}(\underline{x}-\underline{x}_o), \underline{x}-\underline{x}_0) + \delta(\underline{x},\underline{x}_o) \tag{A6-13}$$

where α_o is real, $\underline{B}$ is a positive definite q x q matrix, and

$$\delta(\underline{x},\underline{x}_o) = o(\|\underline{x}-\underline{x}_o\|^2) \quad \text{as} \quad \|\underline{x}-\underline{x}_o\| \to 0 \ . \tag{A6-14}$$

Condition 4:

$$\lim_{\substack{\underline{x}\to\underline{x}_o \\ c\to 0}} E\left((\underline{Z}(\underline{x},-c) - \underline{Z}(\underline{x},c))\ (\underline{Z}(\underline{x},-c) - \underline{Z}(\underline{x},c))^T\right) = \underline{\pi} \tag{A6-15}$$

where $\underline{\pi}$ is a non-negative definite matrix.

Condition 5:

$$\lim_{R\to\infty}\ \sup_{k} \int_{\{\|\underline{Z}_k\| > R\}} \|\underline{Z}_k\|^2\, dP = 0 \tag{A6-16}$$

where

$$\underline{Z}_k = \underline{Z}(\underline{x}_k,-c_k) - \underline{Z}(\underline{x}_k,c_k) \ . \tag{A6-17}$$

Condition 6: There exist positive numbers ε, c_o, and K_1 such that for all

$$c \le c_o \tag{A6-18}$$

and all $\underline{x}$ satisfying

$$c < \|\underline{x}-\underline{x}_o\| < \varepsilon \;, \tag{A6-19}$$

$$\left(\underline{x} - \underline{x}_o,\; (\underline{M}(\underline{x},-c) - \underline{M}(\underline{x},c))\,\frac{1}{c}\right) > K_1\|\underline{x}-\underline{x}_o\|^2 \;. \tag{A6-20}$$

Let $b_1, b_2, \ldots, b_q$ denote the characteristic roots of $\underline{B}$ in decreasing order. Find an orthogonal matrix $\underline{P}$ such that $\underline{P}^{-1}\,\underline{B}\,\underline{P}$ is a diagonal matrix and let

$$\underline{\pi}^* = \underline{P}^{-1}\,\underline{\pi}\,\underline{P} \;. \tag{A6-21}$$

Let

$$w_{ij} = \frac{A^2\;\pi^*_{ij}}{4Ab_i + 4Ab_j - 1} \;, \tag{A6-22}$$

$$\underline{W} = (w_{ij})$$

and define

$$\underline{W}^* = \underline{P}\;\underline{W}\;\underline{P}^{-1} \;. \tag{A6-23}$$

If the six Conditions are satisfied, with

$$K_1 \leq 4b_q \tag{A6-24}$$

in Condition 6, and

$$c_n = n^{-\frac{1}{4}} \tag{A6-25}$$

and

$$a_n = A/n \tag{A6-26}$$

where A is such that

$$AK_1 > \frac{1}{4} \ , \tag{A6-27}$$

$n^{\frac{1}{4}}(\underline{x}-\underline{x}_o)$ is asymptotically normally distributed with mean zero and covariance matrix $\underline{W}^*$.

Blum (23) extended the Kiefer-Wolfowitz method to the multi-dimensional case and proved convergence with probability one. Sacks later determined the asymptotic distribution of the error and it is his sixth theorem (25) stated here.

LITERATURE CITED

1. Carlyle, J. and J. Thomas, "On Nonparametric Signal Detection," *Institute of Radio Engineers (IRE) Transactions on Information Theory,* Vol. IT-10, April 1964, pp. 146-152.

2. Abramson, N. and D. Braverman, "Learning to Recognize Patterns in a Random Environment," *IRE Trans. on Information Theory,* Vol. IT-8, September 1962, pp. 58-63.

3. Fralick, S., *Learning to Recognize Patterns Without a Teacher,* Stanford Electronics Laboratories, Tech. Report No. 6103-10, March 1965, p. 89.

4. Hancock, J. and D. Mix, *Learning Theory Applied to Communications,* Purdue Research Foundation, Tech. Report No. TR-EE-65-20, October 1965.

5. Spragins, J., "Learning Without a Teacher," *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Information Theory,* Vol. IT-12, April 1966, pp. 223-230.

6. Robbins, H. and S. Monro, "A Stochastic Approximation Method," *Annals of Mathematical Statistics,* Vol. 22, 1951, pp. 400-407.

7. Sakrison, D., "Stochastic Approximation: A Recursive Method for Solving Regression Problems," *Advances in Communication Systems,* Ed. by A. V. Balakrishnan, Vol. 2, Academic Press, New York, 1966, pp. 51-106.

8. Kac, M., "A Note on Learning Signal Detection," *IRE Trans. on Information Theory,* Vol. IT-8, February 1962, pp. 126-128.

9. Tong, P. and B. Liu, "Application of Stochastic Approximation Techniques to Two Signal Detection Problems," *Proceedings of the National Electronics Conference,* 1966, pp. 804-809.

10. Cooper, D., *On the Existence of Nonsupervised Adaptive Signal Detectors; and Detector Estimation Using Stochastic Approximation Methods,* Ph.D. Dissertation in Applied Mathematics, Columbia University, 1966, pp. 81-111.

11. Kesten, H., "Accelerated Stochastic Approximation," *Ann. Math. Stat.,* Vol. 29, 1958, pp. 41-59.

12. Nikolic, Z. and K. Fu, "A Mathematical Model of Learning in an Unknown Random Environment," *Proc. of the NEC,* 1966, pp. 607-612.

13. Blum, J., "Approximation Methods which Converge with Probability One," *Ann. Math. Stat.*, Vol. 25, 1954, pp. 382-386.

14. Sacks, J., "Asymptotic Distributions of Stochastic Approximation Procedures," *Ann. Math. Stat.*, Vol. 29, 1958, pp. 373-405.

15. Parzen, E., *Modern Probability Theory and Its Applications*, John Wiley and Sons, Inc., New York, 1960, p. 151n.

16. Cooper, *op. cit.*, pp. 100-103.

17. Marcum, J., "A Statistical Theory of Target Detection by Pulsed Radar," *IRE Trans. on Information Theory*, Vol. IT-6, April 1960, p. 159.

18. Marcum, J., *Table of Q Functions*, Research Memorandum No. RM-339, The Rand Corp., Santa Monica, California, January 1950.

19. Kiefer, J. and J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function," *Ann. Math. Stat.*, Vol. 23, 1952, pp. 462-466.

20. Lebedev, N. N., *Special Functions and Their Applications*, Trans. R. A. Silverman, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1965, p. 110.

21. Sacks, *op. cit.*, pp. 384-390.

22. Blum, J., "Multidimensional Stochastic Approximation Methods," *Ann. Math. Stat.*, Vol. 25, December 1954, pp. 738-740.

23. Blum, *ibid.*, pp. 742-743.

24. Sacks, *op. cit.*, pp. 398-400.

25. Sacks, *op. cit.*, pp. 400-404.

26. Cramer, Harald, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J., 1946, p. 114.

27. Derusso, P. M., R. J. Roy, and C. M. Close, *State Variables for Engineers*, John Wiley and Sons, Inc., New York, 1966, p. 266.

28. Sakrison, *op. cit.*, pp. 71-72.

29. Wozencraft, J. and I. Jacobs, *Principles of Communication Engineering*, John Wiley and Sons, Inc., New York, 1965, p. 83.

VITA

Maurice Joseph Bertrand Bouvier, Jr. was born in Covington, Louisiana, on April 22, 1937, and is the son of Maurice, Sr. and Lucille Burnett Bouvier. On August 24, 1957, he married Carolyn Von Behren of New Orleans and they now have one daughter and three sons.

After graduation from Pensacola High School, Pensacola, Florida, he attended Pensacola Junior College for one semester before transferring to Louisiana State University, Baton Rouge. While an undergraduate, he worked one summer for Gulf Power Company, Pensacola, and during his senior year as a paper grader in the Mathematics Department. Upon receiving the B.S.E.E. in June 1959, he joined the Westinghouse Electric Corporation, and while on their training program was selected for a one-semester graduate course at the University of Pittsburgh. He reported to their Electronics Division, Baltimore, in April, 1960, where he was a member of a radar receiver design group. In May, 1962, he joined the Sperry Microwave Electronics Company, Clearwater, Florida, and subsequently The Boeing Company, New Orleans. He began an evening graduate study program at Louisiana State University in June, 1963, and received the M.S.E.E. in August, 1965.

Since September, 1965, he has been attending the Georgia Institute of Technology with the financial support of a NDEA Fellowship. He is a member of Tau Beta Pi, Eta Kappa Nu, Pi Mu Epsilon and Phi Kappa Phi.